令和２年度博士学位論文

シクラメンの花色生合成に関わる分子生物学的解析

Study on Molecular Mechanism of Flower Coloration in Cyclamen

埼玉工業大学　大学院工学研究科　博士後期課程
生命環境化学　専攻

指導教員　　秦田 勇二 教授

論文提出者　康 暁飛

Contents

List of abbreviation

3GT/UFGT:	Flavonoid 3-*O*-glycosyltransferase
5GT/A5GT:	Anthocyanin 5-*O*-glycosyltransferase
5'RACE:	5' Rapid amplification of cDNA ends
7GT:	Flavonoid 7-*O*-glycosyltransferase
ANS:	Anthocyanin synthase
Amp:	Ampicillin
AT:	Anthocyanin acyltransferase
bp:	Base pair
BSA:	Bovine serum albumin
cDNA:	Complementary deoxyribonucleic acid
C4H:	Cinnamic acid hydroxylase
CHI:	Chalcone isomerase
CHS:	Chalcone synthase
CIA:	Chloroform/isoamyl alcohol (24:1)
CoA:	Coenzyme A
CTAB:	Cetyltrimethylammonium bromide
Cy3G:	Cyanidin 3-glucoside
Cy3,5dG:	Cyanidin 3,5-diglucoside
CYP:	Cytochrome P450
D.W:	Distilled Water
DEPC:	Diethyl pyro carbonate
Dp3G:	Delphinidin 3-glucoside
Dp3,5dG:	Delphinidin 3,5-diglucoside
DFR:	Dihydroflavonol reductase
DHK:	Dihydrokaempferol

DHM:	Dihydromyricetin
DHQ:	Dihydroquercetin
dNTP:	Deoxy-ribonucleotide triphosphate
eEF1α:	*Elongation factor-1α*
EDTA:	Ethylene diamine tetra acetic acid
ER:	Endoplasmic reticulum
F3H:	Flavanone 3-hydroxylase
F3'H:	Flavonoid 3'-hyroxylase
F3'5'H:	Flavonoid 3',5'-hyroxylase
FLS:	Flavonol synthase
FNS:	Flavone synthase
GT:	Glycosyltransferase
GST:	Glutathione transferase
HPLC:	High performance liquid chromatography
IPTG:	Isopropyl-*β*-D-thiogalactopyranoside
LB:	Luria-Bertani medium
Mv3G:	Malvidin 3-glucoside
Mv3,5dG:	Malvidin 3,5-diglucoside
MT:	Methyltransferase
NADPH:	Nicotinamide adenine dinucleotide phosphate
OD:	Optical density
ORF:	Open reading frame
PAL:	Phenylalanine lyase
PCI:	Phenol/Chloroform/Isoamyl alcohol (25:24:1)
PCR:	Polymerase chain reaction
Pg3G:	Pelargonidin 3-glucoside
Pg3,5dG:	Pelargonidin 3,5-diglucoside
Pn3Nh:	Peonidin 3-*O*-neohesperidoside

Pn3G:	Peonidin 3-glucoside
Pn3,5dG:	Peonidin 3,5-diglucoside
Pt3G:	Petunidin 3-glucoside
Pt3,5dG:	Petunidin 3,5-diglucoside
PSPG:	Plant secondary product glycosyltransferase
RNase:	Ribonuclease
SDS:	Sodium dodecyl sulfate
SDS-PAGE:	SDS- polyacrylamide gel electrophoresis
SOC:	Super optimal broth with catabolite repression
SOPMA:	Self-optimized prediction method with alignment
UDP:	Uridine diphosphate
X-Gal:	5-Bromo-4-chloro-3-indolyl β-d-galactopyranoside

Abstract

Cyclamen is one of the world's best-selling potted plants due to better ornamental traits and simple cultivation management. Most of the current ornamental cyclamen cultivars are obtained from a single wild species, purple flower *Cyclamen persicum* (*C. persicum,* 2n = 2x = 48), through natural variation and the hybridization of mutants. These cultivars are always rich in colors, mainly because of an important group of flavonoids, anthocyanin. Investigate the molecular mechanism of cyclamen flower color formation, get the information about genes related to anthocyanin biosynthesis is of great significance for molecular breeding of new varieties with novel colors.

Color mutants are considered to be good materials for studying the function of genes associated with anthocyanin synthesis. *Cyclamen persicum* 'Strass'(STR) is one cultivar of *C. persicum*, which has larger red flowers than the original species. The major anthocyanin component in STR petals has been changed from malvidin 3,5-diglucoside (Mv3,5dG, a major anthocyanin in *C. persicum*) to peonidin 3-*O*-neohesperidoside (Pn3Nh). Meanwhile, the flow of synthesize Pn3Nh and Mv3,5dG is partly controlled by flavonoid 3'-hydroxylase (F3'H) and flavonoid 3',5'-hydroxylase (F3'5'H) respectively. To find out what caused the change in STR flower colors, we isolated the candidate *F3'H* genes, which are highly correlated with color mutant, from STR petals and obtained three open reading frames (ORFs). The amino acid sequences deduced from these genes are highly similar to F3'H protein in other plants, and have conserved motifs of F3'H protein. The results of real-time PCR showed that the transcription level of *F3'Hs* were the highest at the early stage of flower development, and gradually decreased as the flowers bloomed. The expression of *F3'H*s were also detected in leaves. When compared the expression of *F3'H* in STR and *C. persicum*, and the results showed that *F3'H* expressed much stronger in STR

than that in *C. persicum*. All the results suggested *F3'H* is likely taking an active role in pigmentation in STR. We also constructed the recombinant expression vector pET21a-*STRF3'H1*, pET21a-*STRF3'H2a*, pET21a-*STRF3'H2b* and determined the best conditions for protein induction, which provided a basis for analyzing the function of *STRF3'H* gene and identifying the enzyme activity of STRF3'H *in vitro*.

In order to enhance the commercial value of the ornamental cyclamen, researchers are also committed to cultivating cyclamen with various colors and fragrances. The wild species, *Cyclamen purpurascens (C. purpurascens)*, which has a sweet fragrance, had been applied to horticulture breeding of cyclamen. Interestingly, all the F1 progenies of the cross between *C. purpurascens* and *C. persicum* cultivars contain 3,5- diglucoside type anthocyanins in petals, same to *C. purpurascens* (the major pigment is malvidin 3,5-diglucoside) (Takamura et al. 2005). Anthocyanin 5-*O*-glucosyltransferase (A5GT) is responsible for glycosylation at the 5-*O*-position to generate more stable 3,5-diglucoside type anthocyanins. It is indicated that the expression of *A5GT* in petals is dominant, even in the cross of *C. purpurascens* and *C. persicum* cultivars. This time we isolated two complete ORFs of *A5GT* genes from *C. purpurascens*. By analyzing the deduced amino acid sequences, phylogenetic relationships and expression patterns, we concluded that Cpur5GT2 is more likely to encode a typical A5GT, so a prokaryotic expression vector of Cpur5GT2 was constructed and the enzyme assay was performed *in vitro*. The outcomes revealed that *Cpur5GT2* has a valid enzymatic activity for anthocyanin glycosylation and may make positive contribution to cyclamen coloration.

Keywords: Cyclamen; pigment; anthocyanin; flower color; flavonoid 3'-hydroxylase; anthocyanin 5-*O*-glucosyltransferase

Chapter 1 Introduction

Flower is one of the most attractive organs of ornamental plants, and flower color is one of the most direct ornamental characters. Novel flower color can not only enhance the ornamental value of plants, but also bring potential commercial value. The formation of plant flower color mainly depends on three types of plant pigments, carotenoids, flavonoids and betalains. Carotenoids include carotene and xanthophyll, which are fat-soluble pigments that exist in the cytoplasm in a deposited or crystalline state, and usually causing the pigments in the plants to be yellow, orange, or red (Tanaka et al. 2008). Betaines are water-soluble nitrogen-containing pigments that exist in most families of the *Caryophyllales* and a class of fungi. At present, the known betaines can be divided into the yellow etaxanthins and red betacyanins (Strack et al. 2003). Flavonoids are widely distributed in plants and are secondary metabolites of water-soluble aromatics. They are mostly found in cell vacuoles (Tanaka et al. 2008). When the light passes through the petal pigment layer, part of it is absorbed, and part is reflected by the sponge tissue, and then enters people's eyes through the pigment layer, forming the impression of flower color. The appearance of flower colors is closely related to the types and contents of pigments, and is also affected by the pH of vacuoles, types and concentrations of metal ions, co-pigments and environmental factors.

1.1 Anthocyanin

1.1.1 The structure of anthocyanin

Anthocyanin, a group of flavonoids, is widely found in the petals, leaves, pericarp and seed coat of plants, and is one of the important pigments in plants. Anthocyanin is a water-soluble pigment with acidic and basic groups, which is well soluble in more polar solvents.

Anthocyanins exist in different structural forms in different pH environments. Usually, they are stable at low pH, and their stability decreases rapidly at high pH and even degrades. The presence of anthocyanins gives plants a variety of colors, which enriches the ornamental value of plants and also plays pushing role in the reproduction and evolution of species, for examples, attract pollinators and seed dispersers (Huits et al. 1994), protect plants from UV radiation (Bieza et al. 2001), participate in the synthesis of plant hormones (Winkler et al. 1995), against phytopathogens as the phytoalexin (Nicholson et al. 1992; Dixon et al. 1999) and so on.

Anthocyanin is a kind of compounds made by anthocyanidin binding with glycosides. The basic structure is the C6-C3-C6 carbon framework, that is, two aromatic rings (A ring and B ring) are connected by a central three-carbon bridge, which often forms a heterocyclic ring (C ring) (Figure 1-1). Different types of substituents at the 3' and 5' positions determine the type and color of the anthocyanin. The hydroxyl groups at the 3, 5, and 7 position always combine with different types and numbers of glycosyl to form glycosides. The modification of anthocyanin includes hydroxylation, methylation, glycosylation and acylation. In higher plants, pelargonidin, cyanidin and delphinidin are the most basic anthocyanidins, peonidin is formed by methylation of the R1 group of the cyanidin B ring, petunidin and malvidin are formed by the methylation of delphinidin in R1 and R2 groups, respectively. As the number of hydroxyl groups on the anthocyanin B ring increases, the blue tone gradually deepens; as the degree of methylation of hydroxyl group on B ring increases, the blue tone gradually weakens and the red tone gradually strengthens. As the degree of methylation deepens, the redshift becomes more obvious (Winkel-Shirley 2001; Honda et al. 2002). Glycation will cause the anthocyanin color redshift, while acylation will cause the anthocyanin color blue shift (Tanaka et al. 2008).

1.1.2 The biosynthetic pathway of anthocyanin

The anthocyanin biosynthetic metabolic pathway is a branch of the flavonoid pathway and one of the secondary metabolic pathways that have been well studied in higher plants (Holton and Cornish, 1995; Mol et a1., 1998) (Figure 1-2). Phenylalanine is the direct precursor of flavonoid biosynthesis. Anthocyanins are synthesized from phenylalanine in the cytoplasm through a series of enzymatic reactions, followed by different hydroxylation, glycosylation, methylation and acylation, and finally transported to vacuoles for collection. It has been proposed that enzymes related to anthocyanin synthesis form supramolecular complexes through protein-protein interactions and anchor in the endoplasmic reticulum (ER) membrane (Grotewold 2006a). The synthesis of anthocyanin skeleton is mainly divided into three stages: The first stage is from phenylalanine to 4-coumayl CoA, which is regulated by phenylalanine lyase (PAL) and cinnamic acid hydroxylase (C4H), which is also a stage shared by many secondary metabolisms. The second stage is from 4-coumayl CoA and malonyl CoA to dihydroflavonol, which is the key reaction of flavonoid metabolism. This stage provides the precursor substance naringenin for the synthesis of anthocyanin glycosides, auxiliary pigment flavonoids and flavonols. At different sites of naringenin, the hydroxylation reaction can be catalyzed by three different enzymes, flavanone 3-hydroxylase (F3H) catalyzes hydroxylation at C3 position; flavonoid 3'-hyroxylase (F3'H) catalyzes hydroxylation at C3' position, flavonoid 3',5'-hyroxylase (F3'5'H) catalyzes hydroxylation at C3' and C5' position, thus, dihydrokaempferol (DHK), dihydroquercetin (DHQ) and dihydromyricetin (DHM) were produced respectively. The third stage is under the regulation of dihydroflavonol reductase (DFR) and anthocyanin synthase (ANS/LDOX), the colorless dihydroflavonol are converted into leuco-anthocyanins. After the synthesis of the backbone of anthocyanins is completed, the

diversity modification will be carried out. These modifications depend primarily on UDP-glycose-dependent glycosyltransferase (GT), anthocyanin acyltransferase (AT) and methyltransferase (MT). Each site undergoes glycosylation, acylation and methylation modification to form a variety of different anthocyanin types, which make plants show various colors (Winkel-Shirley, 2001). Finally, anthocyanins are transported to the vacuole for storage and play a physiological role with the assistance of vacuolar transport proteins, such as glutathione transferase (GST). The anthocyanin biosynthetic pathway in higher plant can be divided into three synthetic branch pathways: pelargonidin-based pathway (orange red to brick red), cyanidin-based pathway (pink to red) and delphinidin-based pathway (violet to blue). And these three pathways do not necessarily coexist in the same plant, for example, many plants in nature do not have blue flowers, due to the lack of delphinidin branching pathway, such as carnation and rose (Holton and Tanaka 1994; Mol et al. 1999; Yoshida et al. 2009); cymbidium and petunia lack brick red/orange varieties because they do not have pelargonidin-based anthocyanins (Johnson et al. 1999; Forkmann and Heller 1999).

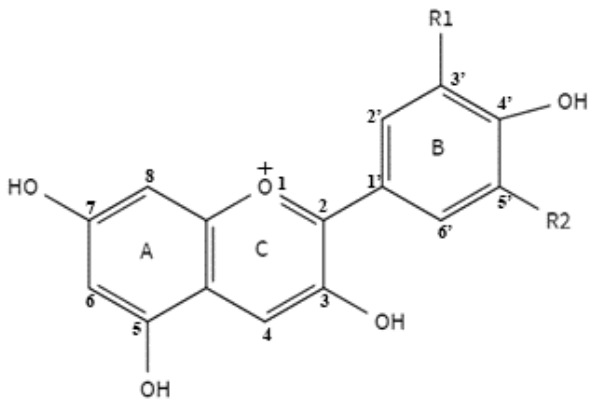


Fig 1-1. The basic structure of anthocyanin.
R1=R2=H pelargonidin; R1=OH, R2=H cyanidin; R1=R2=OH delphinidin; R1=OCH3, R2=H peonidin; R1=OCH3, R2=OH petunidin; R1=R2=OCH3 malvidin

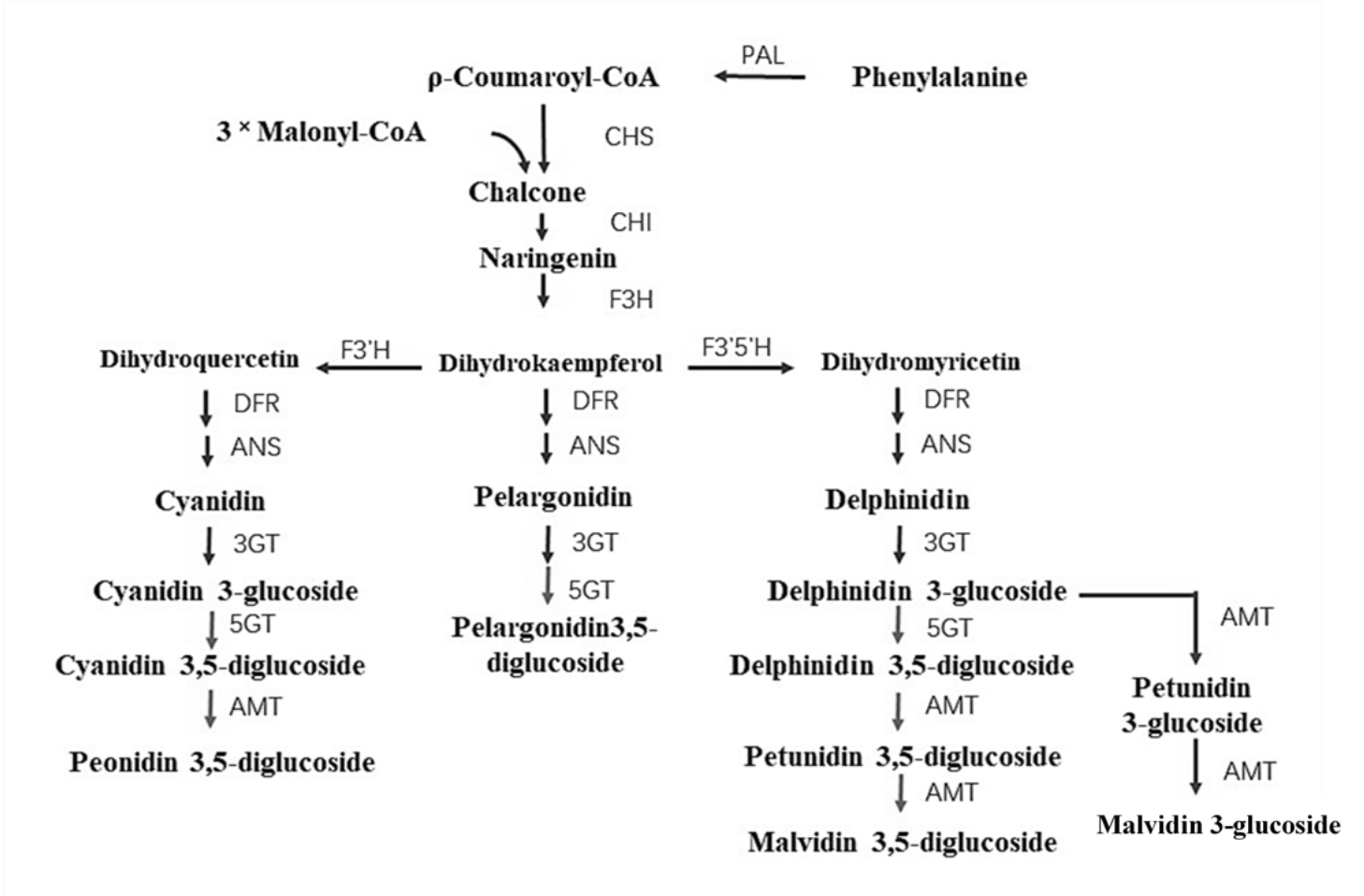


Fig 1-2. The Biosynthetic pathway of anthocyanins in higher plants.

1.2 The main factors related to anthocyanin-based flower color formation

1.2.1 The structural genes involved in anthocyanin biosynthesis

In ornamental plants with anthocyanin as the main pigment, the type and content of anthocyanin are the main factors that determine the final appearance of anthocyanin. Different kinds of anthocyanins show different colors, and when the content of the same anthocyanin changes, the color will change accordingly. For example, when the content of cyanidin gradually increases, the color gradually changes from light red to dark red. This is usually related to the expression of structural genes and/or regulatory factors related to anthocyanin biosynthesis (Nakatsuka et al. 2005).

The first enzyme in the anthocyanin synthesis pathway is CHS. The *CHS* gene was first isolated from parsley by Reimold et al. (1983), which was also the first gene isolated from the flavonoid biosynthetic pathway. Now *CHS* has been isolated from various of ornamental plants, such as arabidopsis (Saslowsky et al. 2000), petunia (Morgret et al. 2005), herbaceous peony (Zhao et al. 2012b). The expression level of *CHS* gene can affect the display of plant flower color. *CHS* gene mutation causes *Torenia fourneiri* flower color to change from blue to white and gray (Fukusaki et al. 2004). Inhibited expression of *CHS* gene is the main cause of the white stripes in petunia 'Red Star' flowers (Koseki et al. 2005).

CHI catalyzes the formation of colorless naringenin. This reaction can proceed spontaneously, but the reaction efficiency can be increased by 7 to 10 times in the presence of CHI (Jez et al. 2000). CHI and CHS play an irreplaceable role in the metabolic branched pathways of flavonoid, flavonol, flavone, procyanidins and anthocyanin synthesis. In *Arabidopsis thaliana*, CHI is a unique enhancer in the flavonoid pathway, and it functions

with TT5 to promote the production of flavonoids (Jiang et al. 2015). When the expression level of *CHI* gene decreases, chalcone and its derivatives will accumulate, and anthocyanin biosynthesis will be hindered. The inactivation of *CHI* from onion (*Allium cepa*) leaded to accumulate of yellow chalcone derivatives, and resulted in a golden mutation in the bulbs of the onion (Kim et al. 2004). Suppressing the expression of *CHI* in tobacco will affect the accumulation of flavonoids in petals and pollen, resulting in color changes (Nishihara et al. 2005).

F3H catalyzes the hydroxylation of the C3 site of flavanones and synthesizes dihydroflavonol, which is considered to be the central point of the anthocyanin biosynthetic pathway. The cDNA of the *F3H* gene was originally cloned from *Antirrhinum majus* (Martin et a1., 1991). Loss of F3H function will hinder the conversion of flavanone to dihydroflavonol, thereby affecting the accumulation of anthocyanins. The mutation of *F3H* (*Transparent Testa 6*) in *Arabidopsis thaliana* leads to a decrease in the pigments in the seed coat, which turns the color of the seed coat gray and white (Peer et al. 2001). Due to the insertion of the retro transposable element TORE1, the *ToreniaF3H* gene was not expressed in white-flowered torenia, which resulted in a decrease in anthocyanin levels in petals (Nishihara et al. 2014).

F3'H and F3'5'H are belonging to cytochrome P450 enzymes (CYP450s) family that regulate hydroxylation of the B-ring. The degree of B-ring hydroxylation has the greatest effect on anthocyanin-based flower color.

DFR catalyzes DHK, DHM and DHQ to produce corresponding colorless leucoanthocyanidin. In addition, these three products (DHK, DHM and DHQ) can further generate flavonols under the catalysis of flavonol synthase (FLS). The competition between FLS and DFR can affect the trend of flavonoid metabolism, thereby changing the flower color. The overexpression of *DFR* increases the anthocyanin content in transgenic tobacco

lines and produces red flower, on the contrary, overexpression of *FLS* promotes the accumulation of flavonols and produces white flowers (Luo et al. 2016). The DFR of some plants has strong substrate specificity, which is one of the reasons why plants show different colors. DFR from *Petunia* cannot catalyze the reduction of DHK to leucopelargonidin, so there is no pelargonidin-based pigment accumulation, and there is no orange-red petunia flower in nature (Forkmann and Ruhnau 1987).

ANS is responsible for catalyzing the oxidation of colorless 1eucoanthocyanidins to produce colored anthocyanidins, which are essential for the formation of plant colors. Mutations in *ANS* genes can also cause anthocyanins to fail to accumulate normally. Gentian *ANS* gene mutation causes its flowers to change from pink to white (Nakatsuka et al. 2005). ANS also affects the synthesis of anthocyanins along with upstream and downstream genes. The flowers of forsythia are yellow because of lack of anthocyanins, and mainly accumulate carotenoid xanthophylls. After transformed *AmDFR* (*Antirrhinum majus*) and *MiANS* (*Matthiola incana*) into forsythia (*Forsythia* × *intermedia* cv 'Spring Glory', the petals of the transgenic plants were detected with carotene as well as cyanidin-based anthocyanidins, and the flower present a novel bronze-orange color (Rosati et al. 2003).

Anthocyanidin is a very unstable substance. Through glycosylation, the unstable anthocyanin is converted into stable anthocyanin, and the maximum absorption spectrum is towards the ultraviolet end. 3GT/UFGT is the last key enzyme in the anthocyanin synthesis pathway. It catalyzes the glycosylation reactions that occur in the anthocyanin synthesis of most plants. Through glycosylation of anthocyanin at 3-*O* position, the solubility of anthocyanin can be improved, which is convenient for transportation and storage to vacuoles. Glucose is the most common glycosylated donor in anthocyanins, others include galactose, rhamnose, xylose, arabinose and fructose. The above types of

sugar groups generally bind to the hydroxyl group at the C3 position of anthocyanins, and sometimes also bind to the hydroxyl group at the C5 and C7 positions. At present, the most researched is 3GT. Kobayashi et al. (2001) found that *3GT* gene was only expressed in red grape variety *Vitis Vinifera*, but not in white grape. Virus-induced silencing of the *RrGT2* gene reduced the accumulation of anthocyanins in the corolla of transgenic tobacco, and tobacco plants showed lighter flowers than normal plants (Sui et al. 2018). Studies on most plants show that the expression level of 3GT is positively correlated with the accumulation of anthocyanins.

MT is one of the key enzymes for anthocyanin modification, and the methylation reaction catalyzed by it contributes to the enrichment of anthocyanin types. Up to date some *OMTs* have be isolated and confirmed as useful molecular tool for altering and diversifying flower color, such as *A3'5'OMT* (Nakamura, 2015), *PsAOMT* (Du et al. 2015), *NmATMs* (Okitsu et al. 2018).

Anthocyanin synthesis in plants is a complex and orderly process involving a variety of structural genes and the enzymes they encode. As early as 1974, scholars put forward the hypothesis of a membrane-associated enzyme complex involved in the metabolism of flavonoids (Stafford, 1974). This multi-enzyme complex greatly increases the rate of enzymatic reactions, and can respond rapidly to signals from inside and outside the cell and change the number and types of terminal products. The preliminary results of studies on the enzymes related to the anthocyanin synthesis of *Arabidopsis* revealed that: Enzymes involved in anthocyanin synthesis assemble into unstable or dynamic linear or spherical complexes in the cytoplasm, weakly anchored on the cytoplasmic side of the ER membrane. CH4, F3'H and other cytochrome P450-dependent monooxygenases act as the "anchor" of the complex fixed to the ER (Burbulis and Winkel-Shirley, 1999; Saslowsky and Winkel-Shirley, 2001; Winkel-Shirley, 2002). The establishment of the multi-enzyme complex for

anthocyanin biosynthesis is helpful to understand the formation of metabolites and the "communication" mechanism in the anthocyanin biosynthetic pathway of higher plants. Of course, the universality of the complex in the entire plant kingdom also requires in-depth exploration.

1.2.2 The transcription factors involved in anthocyanin biosynthesis

The expression intensity and pattern of structural genes related to anthocyanin biosynthesis are regulated by corresponding transcription factors. Transcription factors are a class of proteins that regulate target gene expression through specific binding of DNA sequences and protein-protein interactions to activate or inhibit the transcription of target genes and increase or decrease the level of target gene mRNA. When these transcription factors are mutated, the type or content of anthocyanins will change to varying degrees (Grotewold, 2006). MYB transcription factors are currently considered to be the most critical type of transcription factors in the anthocyanin biosynthesis pathway. One of their main functions is to regulate the synthesis of plant secondary metabolites. MYB transcription factor N-terminal contains a highly conserved DNA binding domain-MYB domain, which generally exists in 1-4 repeats (R). According to the number of R motifs, it can be divided into three types: R1MYB, R2R3MYB and R1R2R3MYB (Jin and Martin, 1999). R2R3-MYB protein plays an important role in the regulation of anthocyanin synthesis. *AaMYB2* (isolated from *Anthurium andraeanum*) abundantly expressed in the spathes from the red, pink, and purple cultivars, but almost cannot be detected in the spathes from the white and green ones. And it also was considered to be involved in regulating the expression of *AaF3H*, *AaANS* (Li et al. 2016). In transgenic Arabidopsis plants, the overexpression of *PsMYB114L* and *PsMYB12L* affected the expression of genes related to anthocyanin synthesis, leading to a significantly higher accumulation of anthocyanins,

resulting in purple-red leaves in Paeonia (Zhang et al. 2019). bHLH transcription factor is named for its basic helix loop helix (basic helix loop helix) conserved domain. The E Box cis-acting element in the promoter region of the gene encoding key enzyme involved in anthocyanin biosynthetic can be specifically recognized by bHLH transcription factors. The transcriptional core region of WD40 protein is called WD40 motif and consists of 40 amino acid residues. WD40 protein is ubiquitous in plants and plays an important role in regulating plant growth and development. The most common mode of action of these transcription factors is to directly bind to the cis-acting elements in the promoter region of the gene to regulate; the other is to combine with each other to form a ternary transcription protein complex and then jointly regulate. For example, MYB-bHLH-WD40 is the most widely studied and discovered complex that regulate the spatiotemporal expression of genes encoding related enzymes in the anthocyanin biosynthesis (Hichri et al. 2011). Nuraini et al. (2020) found in the study of violet color formation in *Matthiola incana*, *MiMYB1* was a key gene that strictly regulated the biosynthesis of anthocyanins in the petals of *Matthiola incana*, and the MiMYB1-MibHLH2-MiWDR1 complex can activate the transcription of endogenous enzyme genes such as *MiF3'H*, *MiDFR* and *MiANS*.

1.2.3 The co-pigment

Some flavonoids and other related compounds combined with anthocyanidins to produce a hyperchromic effect are called co-pigments (Mazza et al. 1993). Co-pigment has no direct effect on flower color, the degree of co-pigmentation was a function of the concentration of the anthocyanins and the molar ratio of co-pigments to anthocyanins (Asen et al. 1972). Flavonols and flavones are common co-pigments, their content will also affect the color. There are two key enzymes flavone synthase (FNS) and flavonol synthase (FLS), which directly affect the synthesis of flavones and flavonol pigments. When *FNS* and/or

FLS genes are mutated, the synthesis of flavones and flavonols is hindered, which may cause flower color changes. There are two different kinds of FNS in the biosynthesis of flavones in plants: FNS I, a soluble dioxygenase, was only described for members of the *Apiaceae* family; FNS II, belong to cytochrome P450 family, has been found in all other flavone accumulating tissues (Martens and Mithöfer, 2005). In transgenic torenia plants with suppressed *FSII* expression, accumulation of both flavone and anthocyanin decreased in the petals, and the flower color changed from blue to pale blue (Ueyama et al. 2002). That implied that flavone contributes to the stability of anthocyanin by co-pigmentation of flavone and anthocyanin. Flower color of Petunia changed from violet to pale violet by the expression of the *FLS* or *FNS* gene (Tsuda et al. 2004). The content of anthocyanin in the petals of transgenic tobacco expressing *OsFLS* was significantly reduced, while the content of kaholinol-3-*O*-rutinoside was significantly increased, and resulting in light pink or white flowers (Park et al. 2019). By regulating the expression of *DFR* and *FLS* genes in tobacco, the accumulation of anthocyanin and flavonol can be influenced to produce red or white flowers (Luo et al. 2016). The heterologous expression of *MaFLS* in tobacco inhibits the expression of *NtDFR*, *NtANS1* and *NtANS2*, resulting in reduced petal coloration (Liu et al. 2019).

1.2.4 Other factors

Anthocyanins generally exist in petal epidermal cells, so the shape of epidermal cells affects the appearance of flower color to a certain extent. The cone-shaped cells can increase the proportion of incident light entering the epidermal cells and enhance the light absorption of the pigment, thereby enhancing the intensity of the flower color. Noda et al. (1972) has identified a gene *(mixta)* that affects the intensity of the epidermal cell pigmentation in *Antirrhinum majus* petals by affecting the shape of petal epidermal cells.

Anthocyanins are mainly stored in vacuoles of petal epidermal cells, and the pH value of vacuoles also affects the color presentation of anthocyanins. Due to the mutation and inactivation of the *Pr* gene encoding InNHX1 (vacuolar Na^+/H^+ exchanger) in Japanese morning glory, the vacuole PH increases during the flower-opening, and finally its reddish-purple buds changed into purple flowers (Yamaguchi et al. 2001). In addition, some metal ions could combine with the *O*-dihydroxyl position of the anthocyanin B ring to form highly colored and stable metal complexes. Sigurdson et al. (2017) complexed different types of cyanidin derivatives with Fe^{3+}, Al^{3+}, and found that both ions can make anthocyanins produce a certain red shift and color enhancement, but $Fe3^+$ has the most obvious effect.

In addition, anthocyanin synthesis is the result of both internal and external factors. Enzymes encoded by structural genes and regulator genes determined the type of anthocyanin, and environmental factors can not only affect the rate of anthocyanin biosynthetic, but also affect the accumulation and stability of anthocyanin. When anthocyanin is accumulated in plant cells, environmental factors have an effect on its stability, thereby accelerating or slowing down the degradation of anthocyanin.

1.3 Research progress on molecular breeding of cyclamen flower color

Cyclamen is a member of *Primulaceae*, native to the Mediterranean coast of Europe, the genus consists of 22 species (Grey-Wilson, 2002). It was once called *kuklos* (means “circle, wheel”) in Greek, mainly due to its spherical tuber and the spiraled peduncle (Cornea-Cipcigan, 2019). The swept-back petals of cyclamen are elegant and beautiful, the leaves also diverse in shapes and rich in patterns. Because of these excellent ornamental characteristics, cyclamen is loved by people all over the world. In recent years, with the expansion of market demand, cyclamen producers and breeding experts have begun to pay

more attention to consumers' preferences and commit to cultivating new varieties with rich and unique ornamental characteristics and high quality. Hybrid breeding is a traditional and important means to create new varieties or types of flowers. It can not only combine the excellent characteristics of two or more varieties, but also produce huge heterosis, improve the growth vigor and stress resistance of flowers. However, there are also some defects in hybrid breeding, such as incompatibility of distant hybridization, difficulty in breaking plant reproductive isolation, long breeding cycle and inconveniency for improvement of single traits. Mutation breeding, which has a relatively short history, is usually performed by radiation mutagenesis and is often combined with tissue culture techniques. It has the advantages of high mutation rate, obtaining more excellent variation types in a short time, etc. But at the same time there are disadvantages such as difficult to grasp the direction of inducing mutation and less favorable variation. With the continuous development of biotechnology, genetic engineering methods have become complementary methods of conventional breeding. It breaks the boundaries of communication between species, provides great potential for improving and modifying flower traits, and provides technical guarantee for directional breeding of flowers. Genetic engineering breeding has also attracted the attention of many cyclamen researchers, especially on the improvement of main ornamental traits.

Most of the current commercial cyclamen are obtained through crossing among selected natural mutants of wild *Cyclamen persicum* Miller. The wild *Cyclamen persicum* always has small flowers consisting of a deep purple "eye" (the base region of the petal) and a purple "slip" (the region excluding the eye) (Grey-Wilson, 2002). However, no matter what color the slips of *Cyclamen persicum*-derived cultivars are, they almost have the same "eye" as the wild *C. persicum*. These cultivars are widely used in potted plants and garden planting because of rich colors and graceful posture. The flower color formation of

cyclamen is most closely related to anthocyanins, an important class of flavonoids, except for a few yellow varieties whose main component is chalcone. The cultivar 'Golden Boy', lacking anthocyanins, has a pale-yellow flower with chalcone 2'-glucoside as a major pigment. 'Pure White', lacking anthocyanins, has a white flower with quercetin and kaempferol glycosides (Ishizaka, 2018). Therefore, the use of molecular technology to regulate the synthesis of anthocyanins to create more new colors of cyclamen seems to be a direct and effective choice.

With the discovery of anthocyanin biosynthesis pathway in plants, some genes related to cyclamen anthocyanin synthesis have been cloned. Antisense inhibits the expression of endogenous *F3'5'H*, lead to a shift in hue from purple to red/pink in one transgenic line (Boase et al. 2010). HPLC analysis showed that the content of delphinidin-derived pigments was reduced while cyanidin-derived pigments was increased. Total anthocyanin concentration was reduced while flavonol concentration was recorded slightly. The results indicated that the regulation of key enzymes can change the direction of different branches of the anthocyanin synthesis pathway, and even affect the content of anthocyanins and other flavonoids. With the development of hybrid breeding and the application of ion beam irradiation technology, more and more cyclamen with novel flower color have been cultivated. Comparative analysis using these color mutations is an effective approach to study the function of genes involved in anthocyanin synthesis. By comparing the content and types of anthocyanin and flavonol in *Cyclamen graecum* gra6 (pink-purple slip and deep purple eye) and *C. graecum* gra50 (white-flowered), it was found that the difference of color between them was probably due to the interruption of anthocyanin synthesis. Therefore, the key genes involved in their anthocyanin synthesis were isolated and expressed, and the expression defective of *CgraDFR2* gene was probably found to be the main cause of white flower mutation (Akita et al. 2010). By using the ion beam irradiation,

the variant KMrp (the major anthocyanin is delphinidin 3,5-diglucoside) was obtained from KM (the major anthocyanin is malvidin 3,5-diglucoside), and OMT is the key enzyme that catalyzes the synthesis of malvidin 3,5-diglucoside from delphinidin 3,5-diglucoside. In view of this, two *CkmOMT* genes were isolated, and it was proved that a deletion of the entire *CkmOMT2* region caused by ion-beam irradiation led to its defective expression in KMrp, thus the change in anthocyanin composition in KMrp (Akita et al. 2011). The biosynthesis of anthocyanins is carried out in the cytosol, and then must be transported to the vacuole to make the plant tissues show colorful colors. A *GST* gene related to vacuolar accumulation of anthocyanins in cyclamen was be identified, which will provide further insight into the synthesis and transport of anthocyanin (Kitamura et al. 2012). Flavonols known as co-pigment that occasionally modify flower color when combined with anthocyanins. The functional *FLS* genes have been isolated from *Cyclamen purpurascens,* which greatly broadened our understanding of cyclamen flower coloration (Akita et al. 2018). Cloning and characterization of these key genes helped us gradually understand the molecular mechanism of cyclamen anthocyanin synthesis. And through the development of molecular technology, these genes will also become beneficial tools to change the content and type of anthocyanins, thereby changing the appearance of flower colors.

Chapter 2 Isolation and analysis of f*lavonoid 3'-hydroxylase* (*F3'H*) genes involved in flower coloration from Cyclamen

2.1 Introduction

Among the structural genes involved in anthocyanin biosynthesis, *F3'H* and *F3'5'H* played crucial roles in determining the hydroxylation pattern of flavonoids (Figure 2-1). F3'H catalyzes the hydroxylation at the C3 position of the B ring to generate DHQ, which is then catalyzed by a series of downstream enzymes to finally generate cyanidin-based red anthocyanin. F3'5'H catalyzes the hydroxylation at the C3, C5 position to produce DHM, and finally generate delphinidin-based bule anthocyanin. In addition, F3'H usually exhibits broad substrate specificity and can also react on multiple sites such as flavone, flavonol, flavanol, dihydroflavonol, etc. For example, it can catalyze kaempferol to quercetin and naringenin to eriodictyol (Forkmann, 1999). Both F3'H and F3'5'H belong to cytochrome P450 family, and they are the member of CYP75B and CYP75A subfamily respectively. They mainly catalyze hydroxylation and require NADPH as a cofactor (Forkmann, 1991). Since the F3'H gene was first cloned from petunia in 1999, the *F3'H* gene has now been isolated from a variety of plants, such as snapdragon, tulip, gentian, gerbera and so on. F3'H is a key enzyme in the flow of cyanidin-based anthocyanin synthesis, so it is generally believed that the expression of *F3'H* is related to the accumulation of red anthocyanins. Down-regulation of the expression of *F3'H* would change the proportion of cyanidin-type anthocyanin in flowers and lead to the change of flower color. Down-regulation of endogenous *F3'H* expression and over-expression of rose *DFR* gene in red petunia resulted in the promotion of pelargonidin synthesis to produce orange petunia (Tsuda, 2004). Han et al. (2010) detected the accumulation of cyanidin-type and pelargonidin-type anthocyanins in *Arabidopsis* seedlings transformed with apple *MdF3'H* gene however,

these anthocyanins were not detected in *Arabidopsis tt7-1* mutant seedlings. Expressing the *F3'H* gene of *Snapdragon Antirrhinum* kelloggii in petunia increases the cyanidin content in the transgenic plants and the flower color becomes redder (Ishiguro, 2012). The expression of *F3'H* in tulips is positively correlated with the accumulation of cyanidin. However, after the insertion mutation of *F3'H* promoter, its transcriptional activity decreased, which hindered the synthesis of cyanidin in petals, resulting in the production of light-colored varieties (Yuan Y. et al. 2014). *F3'H* of *Euphorbia pulcherrima* cultivar 'Harvest Orange' contained an insertion of 28 bases, which caused frameshift mutation with a premature stop codon, resulting in nonfunctional enzymes. In the absence of F3'H enzyme activity, the plants have enough pelargonidin precursor to give bracts of 'Harvest Orange' rare orange-red color (Nitarska et al,2018). These indicate that regulating the expression of *F3'H* can indeed regulate the accumulation of anthocyanins in plants, thereby changing the color of plants. But up to now the molecular and biochemical characterization of cyclamen F3'H has almost not been described.

STR is one cultivar of *C. persicum*, which has larger red flowers than the original species. Component analysis displayed that the major anthocyanin components in STR petals has been changed from Mv3,5dG to Pn3Nh. F3'H and F3'5'H are key enzymes for synthesizing Pn3Nh and Mv3,5dG respectively. This time we isolated *F3'H* genes from STR slips obtained three full-length ORFs, analyzed the expression patterns of *F3'Hs*, constructed the corresponding protein expression vectors and determined the best induction conditions, which laid a foundation for further study on the relationship between F3'H and flower color formation in the future.

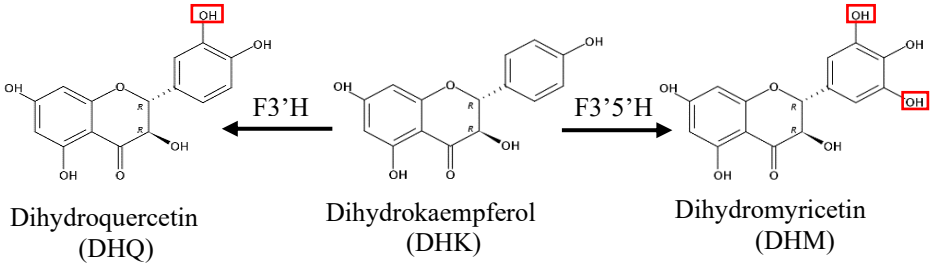
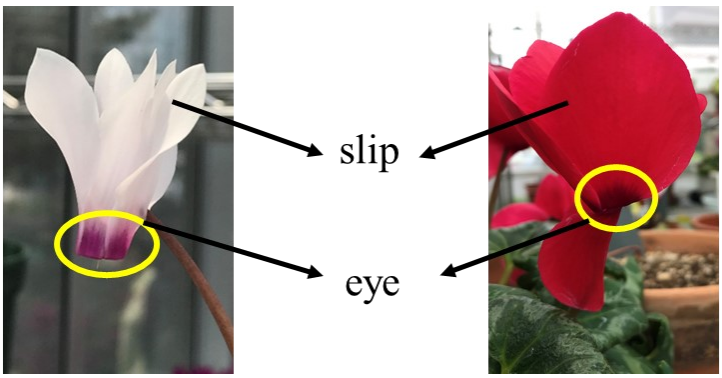


Fig. 2-1 Schematic diagram of the catalytic reactions of F3'H and F3'5'H in the anthocyanin synthesis pathway

2.2 Materials and methods

2.2.1 Plant materials

C. persicum and STR were grown in greenhouse facility at Saitama Institute of Technology. The petals of cyclamen were divided into two parts: a base part known as the "eye", and all other part of the petal called "slip" (Figure 2-2). Leaf and slips were sampled and immediately frozen in liquid nitrogen then kept at -80 °C until required.



(a) *C. persicum* (b) 'Strauss'

Fig. 2-2 The photo of plant materials (a) *C. persicum*, (b) 'Strauss'
The circled parts are called "eye", and the rest parts of the petal called "slip".

2.2.2 Characterization of anthocyanidin composition

Taking full opened slips of STR as materials, 10% acetic acid by 10 times the volume of fresh weight was added and grinded thoroughly in a mortar, transfer to a centrifuge tube, and centrifuge, 13200 rpm, 15 min, 4 ℃. Aspirate the supernatant, filter it with a 0.22 μm microporous membrane, and store in a refrigerator at 4 ℃, shaded from light. High-performance liquid chromatography (HPLC) was performed. A Prodigy ODS-3 reversed-phase column (4.6 × 100 mm 3 μm 100 Å, Phenomenex) was used to separate the metabolites at 30 ℃. The mobile phase consisted of 1.5% (v/v) phosphoric acid(A), 1.5% (v/v) phosphoric acid, 20% (v/v) acetic acid and 25% (v/v) acetonitrile solution (B). The elution program was proceeded over 60 min at a flow rate of 0.3 ml min^{-1}. Quantify the reaction products by measuring the absorbance peak area at 530 nm.

2.2.3 Extraction of genomic DNA

The method of DNA extraction was modified from Murray and Thompson (1980). Plant material was ground to powder in liquid nitrogen. Add 500 μl of 2% Cetyltrimethylammonium Bromide (CTAB) solution per 100 mg materials. Recovered the mixture to a 1.5 ml centrifuge tube, incubated at 65 ℃ for 30 min. Put in 200 μl of chloroform/isoamyl alcohol (24:1) (CIA), after centrifugation (15000 rpm, 15 min), pipette the supernatant into a new centrifuge tube. Add 200 μl of CIA and centrifuge again, collect the supernatant, mixed with 1.5 times volume of 1% CTAB leave at room temperature for 1 hour. Centrifuge (8000 rpm, 10 min), then discard the supernatant and add 400 μl of CsCl until the precipitate is completely dissolved. Continue to add 100% alcohol to the precipitation solution, mix thoroughly, and place it at -20 ℃ for more than 20 minutes. Centrifuge (15000 rpm, 15 min, 4 ℃) after placing, discard the supernatant, add 70% alcohol for elution (15000 rpm, 5 min, 4 ℃). Removed residual ethanol by drying in a

Vacuum drying. Use 45 μl of TE to dissolve the precipitate, add 5 μl of RNase A (5mg ml^{-1}), and place at 37 °C for 1 hour to remove RNA interference.

After the RNase treatment, add equal volumes of Phenol/Chloroform/Isoamyl alcohol (25:24:1) (PCI) (15000 rpm, 5 min, 4°C) and CIA (15000 rpm, 5 min, 4°C) for extraction, and finally recover the supernatant to a new centrifuge tube. Precipitate DNA with 2.5 times volume of 100% alcohol and 1/10 volume of 3 M sodium acetate (NaAc) (15000 rpm, 15 min, 4°C), keep the precipitate, add 500 μl of 70% alcohol for elution (15000 rpm, 5 min, 4°C). Discard the supernatant, dry the precipitate, add an appropriate amount of TE solution to dissolve the precipitate, and finally store at -20 °C.

2.2.4 Extraction of total RNA

Total RNA was extracted from leaves and the slips at four different floral development stages: 1) bud between 0.2 cm and 0.5 cm; 2) bud between 0.6 cm and 1.0 cm; 3) bud more than 2.0 cm long; 4) full opened flower (Figure. 2-3). The method modified from CTAB method (Chang et al. 1993).

Weigh 100mg of plant tissue, add liquid nitrogen and quickly grind in a mortar to powder. For each100 mg homogenized tissue of plant, used 1000 mL of 2×CTAB solution, 40 μl of 4% *β*-mercaptoethanol. Transferred the homogenate to the microcentrifuge tube and warmed in heat-block at 65 °C, 10 min. Added 200 μl of CIA, stirred then centrifuged the sample for 5 min,13 000 rpm, 20 °C, to separate the phases. Transferred the aqueous supper phase to a new tube. Added 200 μl of CIA and centrifuged again. Added a quarter of 10 M LiCl and incubate on ice for at least 2 hours. Centrifuged the sample for15 min, 15 000 rpm, 4 °C, then removed the aqueous supper phase. Dissolved sample in100 μl of TE (10 mM Tris-HCI, 1mM EDTA pH 8), centrifuged for 15 min at 13 000 rpm, 4 °C. Transferred the aqueous supper phase to a new tube. Ethanol precipitation of RNA: add 3

M sodium acetate and 2.5 times the volume of 100% Ethanol, centrifuged for 13 000 rpm 40 ℃. Decanted the super natant without disturbing the pellet and subsequently wash with 500 µl 70% ethanol then centrifuge at 15 000 rpm, 4 ℃ for 5 minutes. Removed residual ethanol by drying in a vacuum drying. DNase treatment: added 43 µl of RNase-free water, 5 µl of 10×DNaseI Buffer and 2 µl of DNase I (0.2 units/µl), then incubated for 1 hour at 37 ℃ in water bath. Added volumes equal of PCI to RNA sample and centrifuged for 5 min at 15 000 rpm, 4 ℃. Transferred the aqueous upper phase to new microcentrifuge tube, added volumes of equal of CIA to RNA sample and centrifuged for 5 min,15 000rpm, 4 ℃. Ethanol precipitation of RNA: add 3 M sodium acetate solution, 100% Ethanol, centrifuge for 13 000 rpm 4 ℃. Decanted the super natant without disturbing the pellet and subsequently wash with 500 µl 70 % ethanol, then centrifuge at 15 000rpm 4 ℃ for 5min. Removed residual ethanol by drying in vacuum drying. Added 50 µl RNase-free water to dissolve. Store at -80 ℃ until use.

Fig. 2-3 Different floral development stages of STR

2.2.5 Synthesis of first-strand cDNA

First-strand cDNAs were synthesized from 2 μg total RNA extracted from slips of all stages using an oligo (dT)-anchor primer (5'-GAC TCG AGT CGA CAT CGA T_{17}-3') with reverse transcriptase according to the manufacturer's instructions (PrimeScript II 1st cDNA synthesis kit, TaKaRa, Japan). Store the synthesized cDNA at -20 ℃.

2.2.6 Isolation of **STRF3'H** *genes*

The degenerate primers were designed based on the conserved domains of plant F3'H proteins involved in anthocyanin accumulation, such as *Arabidopsis thaliana* (*AtTT7*, AF155171, 2000), *Petunia* (*PhF3'H*, AF155332, 1999), *Centaurea cyanus* (*CcF3'H*, FJ753550,2009) and *Vitis amurensis* (*VaF3'H*, FJ645766, 2009). The partial cDNA of the *F3'H* homologues were isolated from slips of STR by reverse transcription-polymerase chain reaction (RT-PCR). To isolate the putative *STRF3'H* full-length cDNA, 3'rapid amplification of cDNA ends (RACE) method (Frohman et al. 1988) and 5'-RACE method were carried out by using a 5'/3'-RACE 2nd Generation Kit (Roche, Germany). Four pairs of gene special primers (Table 1.) were designed to amplify the 5' cDNA ends of *STRF3'Hs*. To obtain the full-length open reading frames, the cDNA templates were amplified by PCR with the gene-specific primers list in Table 1.

Recovery of PCR products: Gel electrophoresis for PCR products at 220 V for about 20 minutes. Put the gel after electrophoresis under UV light, cut and recover the target band part of the gel as soon as possible. The recovery operation is in accordance with the kit (NucleoSpin® Gel and PCR Clean-up, MACHEREY-NAGEL, Germany).

Ligation: All PCR products were cloned into the pTAC-2 Easy vector (Bio Dynamics Laboratory Inc., Japan). The preparation of ligation reaction system follows the kit instructions, and incubated at 16 ℃ for 30 min.

Desalting: Add 20 μl of PCI to the ligation product, mix and centrifuge (15 000 rpm, 4 °C, 5 min). Aspirate the supernatant and transfer to a new 1.5 ml tube, add 20 μl of CIA, and centrifuge (15,000 rpm, 4 °C, 5 min). Transfer the supernatant to a new 1.5 ml tube, followed Ethanol precipitation and ethanol rinsing. After drying in a vacuum desiccator for 5 min, add an appropriate amount of TE buffer (PH8.0) to dissolve the precipitate.

Electroporation: Thawed 40 μl of competent cells (*Escherichia coli* JM109) on ice and added 2 μl of the desalted ligation products, mixed and transferred the mixture to gap of 0.1 cm electrode cuvette. Transformation was performed by electroporation (Gene Pulser Xcell ™), selected the protocol detail screen for *E. coli* to pulse 1.8 kV. Add 700 ml SOC medium, repeatedly pipette and aspirate, then transfer all the bacterial liquid to a new centrifuge tube, and incubated for 1 hour in a 37°C water bath. Spread the incubated bacterial solution with 40 μl of 5-Bromo-4-chloro-3-indolyl *β*-d-galactopyranoside (X-gal) (20 mg L^{-1}) and 4 μl of isopropyl-*β*-D-thiogalactopyranoside (IPTG) (200 mg mL^{-1}) on the LB medium containing 100 mg ml^{-1} of Ampicillin (Amp). Place in a 37 °C incubator for 16 hours.

Colony PCR: Take the detection of 8 samples as an example, and all operations were performed on ice. Prepare the reaction solution according to the following system: 16 μl of 1×Standard Reaction Buffer, 16 μl of 2 mM dNTP, 1.6 of 1mM primer (T7-Fw, SP6-Rv) respectively, 0.64 μl of DNA Taq Polymerase, vortex to mix, dispense 20 μl of the mixture to 8-strip PCR tubes on ice. Take another 8-strip PCR tube and inject 50 μl LB liquid medium into each tube. Use a sterilized toothpick to pick out a single white clone, quickly dip it into the small tube containing the PCR reaction solution, and then dip it into the corresponding small tube containing the LB culture solution. Follow this method to pick 8 white clones in turn. Store the 8-strip PCR tubes containing LB medium at 4 °C. Put the 8-strip PCR tubes containing the PCR reaction solution into the Thermal cycler. The PCR

reaction program was set as follow: step1，95 ℃ 2 min; step2，95 ℃ 30 s，48 ℃ 30 s, 72℃ 1 min，30cycles; step3, 4 ℃. PCR products were detected by electrophoresis. Take a new test tube and inject 1.5 ml of LB medium containing 100 mg ml^{-1} of Amp. Transfer 50 ml of LB culture solution containing colony corresponding to the reaction solution with the target band into the new test tube, and culture with shaking at 37 °C, 130 rpm, 16 h.

Plasmid extraction (Small-scale): Lysis of Cells, pour the culture into a 1.5 ml microcentrifuge tube. Centrifuge at 8000 rpm, 5 min. After that, remove the medium by aspiration, leaving the bacterial pellet. Add 150 μl Sol.1 (25 mM Tris-HCl, 10 mM EDTA, 50 mM glucose) to resuspend the bacterial pellet. Add Sol.2 (0.2 N NaOH, 1.0% SDS) and Sol.3 (10% HAc,5 mM NaAc) in turn, and mix the centrifuge tube upside down after each solution is added. Add 200 μl of CIA, mix the organic and aqueous phases by vortex, and then centrifuge at 15000 rpm for 5 min at 4 ℃, then transfer the aqueous upper layer to a new tube. Recovery of Plasmid DNA, add 600 μl of isopropanol and mix, place the mixture at room temperature for 10 min. Centrifuge at 15000 rpm, 15 min 4 ℃, abandon supernatant, add 700 μl of 70% EtOH, centrifuge 15000 rpm, 5 min 4 ℃, discard the supernatant and dry the sediment. Dissolve the nucleic acids in 45 μl of TE (pH 8.0), add 5 μl RNase A (5 mg ml^{-1}), vortex the solution gently, 37 ℃ water bath treatment for 1 h. Added 50 μl of PCI to the mixture and centrifuged for 5 min at 15 000 rpm, 4 ℃, transferred the aqueous upper phase to new tube, added 50 μl of CIA and centrifuged for 5 min, 15 000 rpm, 4 ℃. Recover supernatant to new 1.5 ml tube, ethanol precipitation and ethanol rinse, pour away the alcohol and retain the sediment removed residual ethanol by drying in vacuum drying. Added 100 μl of polyethylene glycol (PEG) solution, after the precipitation is dissolved, let it stand at 4 ℃ for 1 h. Then centrifuge (10000 rpm, 10 min, 4 ℃), discard the supernatant and add 200 μl of 70% EtOH, centrifuge (15000 rpm, 5 min, 4 ℃), pour away the alcohol and retain and dry the sediment in vacuum drying. Add 50 μl of TE buffer

(PH 8.0), store at -20 ℃ until use.

Sequencing: Sequenced by a DNA sequencer (Model 3500, Applied Biosystems, USA) using the Big Dye® Terminator ver. 3.1 Cycle Sequencing Kit (Applied Biosystems, MA, USA).

2.2.7 Bioinformatics analysis of STRF3'H sequences

Predict the physical and chemical properties and other information by online software ProtParam (http://web.expasy.org/protparam/). Transmembrane domain was predicted by TMHMM (http://www.cbs.dtu.dk/services/TMHMM/). Prediction of protein secondary structure was carried out by Network Protein Sequence Analysis (Combet et al. 2000) using self-optimized prediction method with alignment (SOPMA) (Geourjon and Deléage 1995). Multi-alignment analysis was performed by the ClustalW (Thompson et al. 1994) program, the deduced amino acid sequence of *STRF3'H1* and *STRF3'H2a*, *STRF3'H2b* were aligned with other F3'H proteins that obtained from the DDBJ/GenBank DNA databases (*TT7*, AF155171; *PhF3'H*, AF155332; *VaF3'H*, FJ645766). Phylogenetic trees were constructed using the Neighbor-Joining method (Saitou et al. 1987) with MEGA7 (Kumar et al. 2016).

2.2.8 Genomic PCR of STRF3'Hs

Based on the obtained full-length *STRF3'H* cDNAs, four pairs of gene-specific primers were used to analyze the genomic sequences in detail. Cycling conditions were as follows: pre-denaturation at 95 ℃ for 2 min, followed by 30 cycles of amplification (95 ℃ 30 seconds, 59-65 ℃, dependent on the melting temperature of each primer, for 30 seconds and 72 ℃ for 2 min), then extension at 72 ℃ for 7 min, at last cooling to 4 ℃. All PCR products were cloned and sequenced using the same method as 2.2.6.

2.2.9 Amplification of STRF3'5'H *and corresponding genomic sequence*

STRF3'5'H was cloned from *C. persicum* (GQ891056). To amplify the corresponding genomic sequence of *STRF3'5'H*, two pairs of gene-specific primers were used (table 1). For genomic PCR of *STRF3'5'H*, 0.5 μg DNA extracted from the leaves of STR were used as template, the cycling conditions were as followed: pre-denaturation at 95 ℃ for 2 min, followed by 30 cycles of amplification (30 seconds at 95 ℃, 30 seconds at 60 ℃ and 2 min at 72 ℃), then extension at 72 ℃ for 7 min, at last cooling to 4 ℃. The PCR product was recovered, ligated to the cloning vector and sequenced; the methods were as 2.2.6 described.

2.2.10 Expression pattern analysis of STRF3'H *genes*

The expression level of *STRF3'H* genes at different flowering stages as well as the expression level in leaves were detected by real-time PCR. Real-time PCR was conducted on the Quant Studio™ 1 System using the standard cycling mode with PowerUp™ SYBR® Green Master Mix (Thermo Fisher Scientific). Each 20 μl reaction contained 100 ng of cDNA (template), 10 μM of primer, and 10 μl of 2× Master Mix. The *eEF1a* genes of STR were amplified as an internal control under the same conditions. Each experiment was done in duplicate and repeated at least twice, $2^{-\Delta\Delta CT}$ method was used to analyze the relative expression level of target genes (Livak and Schmittgen 2001).

The transcription level of *F3'H* genes and *F3'5'H* in STR and *C. persicum* were also analyzed by real-time PCR. The detection was performed on the Eco™ Real-Time System with Brilliant III Ultra-Fast SYBR® Green QPCR Master Mix (Agilent Technologies), under the following cycling conditions: 95 °C for 60 s, followed by 40 cycles of 95 °C for 5 s and 60 °C for 60 s. The primers were present in table 1. The *eEF1a* genes of STR and *C. persicum* were amplified as an internal control under the same conditions. Each

experiment was conducted in four replicates, and relative gene expression was determined using the $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen 2001).

2.2.11 Expression of STRF3'Hs *in* E. coli

Take the cDNA of *F3 H1*, *F3'H2a*, *F3'H2b* as the template respectively and use the gene-specific primer with the restriction site (forward with BamHI site and reverse with SacI site) for PCR amplification. Subcloned the amplified fragments into the pTAC-2 vector (Takara), and sequenced to confirm that the sequences were correct. Excised these target fragments by BamHI/ SacI double digestion and ligated in to pET21a (+) vector (Novagen) digested in advance by the same enzymes. Ligation system according to the ligation high (Toyobo, Japan) instructions. Transform the ligation mixture into *E. coli* DH5α (Nippon Gene, Japan) competent cell by heat shock method, and spread the bacteria with recombinant plasmid on LB medium containing AMP (100 mg ml^{-1}). After the plaque grows, pick a single clone to extract the plasmid and sequence. The operation method is as described above. Transfer the correctly sequenced recombinant plasmid, pET21a-*STRF3'H1*, pET21a-*STRF3'H2a*, pET21a-*STRF3'H2b* to *E. coli* BL21(DE3) (Novagen). Take pET21a-*STRF3'H1* as an example:

Pre-culture: Inoculate single colonies to 2 ml LB broth with 100 mg ml^{-1} AMP incubated at 37 °C 16 hours. Take 20 μl of the cultured bacteria liquid and add it to 2 ml LB liquid medium containing AMP (100 mg ml^{-1}), and cultivate at 37 °C until OD600 reached 0.4-0.6.

Main culture: Add IPTG to a final concentration of 100 μM, 500 μM, respectively, and further culture the cells at 20 °C, adjust the incubation time for10 h, 16 h, 20 h.

Separation of soluble and insoluble proteins: Take the cultured bacteria liquid 1.3 ml to 1.5 ml tube, centrifuge at 4000 rpm, 4 °C, 10 min, discard the supernatant, add the

remaining 700 μl bacterial solution to the centrifuge tube, and suspend the precipitate. Add 70 μl of Fast Break™ Cell Lysis Reagent (Promega), mix by inversion for 40 minutes, then centrifuge at 132 000 rpm, 4 °C, 15 min, transfer the supernatant to a new 1.5 ml tube, and store it at 4 °C together with the tube containing the precipitate.

Purification of soluble protein: the fusion protein was purified utilizing the His Link™ Spin Protein Purification System (Promega) following the kit instruction.

SDS-PAGE: Add 50 μl of 10 mM-Tris-HCl buffer solution to the insoluble protein and stir until the precipitate is dissolved. Take 15 μl of insoluble protein, soluble protein, and purified protein, respectively, mix with 5 μl of loading buffer to SDS-PAGE electrophoresis. After completion of electrophoresis, SDS-PAGE electrophoretic gel was stained with Coomassie Brilliant Blue R250 and eluted with eluent (25 ml methyl alcohol, 35 ml acetic acid, Milli-Q water up to 500 ml).

Mass cultivation and expression: Inoculate single colonies of *E. coli* BL21(DE3) harboring pET21a-*STRF3'H1* plasmid to 2 ml LB broth with 100 mg ml^{-1} AMP incubated at 37 °C 16 hours. Take 500 μl of the cultured bacteria liquid and add it to 50 ml LB AMP (100 mg ml^{-1}) liquid medium, and cultivate at 37 °C until OD600 reached 0.4-0.6. Add IPTG to a final concentration of 500 μM further cultured at 20 °C for 20 h. Transfer the cultured bacterial solution to a 50ml falcon tube, centrifuge at 3000 xg,15 min, 4 °C, discard the supernatant, and add 1/20 volume of the bacterial solution Lysis buffer to quickly suspend the precipitation. Add 1/50 volume of the bacterial solution Lysozyme solution, stir on ice for 30 minutes, then add 1/50 volume of the bacterial solution 10% TritonX-100 solution, continue stirring on ice for 20 minutes, centrifuge for 30 minutes (12 000 xg, 4 °C), and recover the supernatant.

Table 1 . Primers used in chapter 2.

Primer	Sequence (5'-3')
For 5'RACE	
C.perF3'H1-RP	TTAAGCCTGGTAAACTTCCTTAGCG
C.perF3'H1-RP1	ATCAGCATTGGGCCTTTCG
C.perF3'H1-RP3	CTCCTTGATGATGGCTTGTAGG
STRF3'H1-RP5	CCGTGTTTCTGGTCACTTCC
STRF3'H1-RP6	CCAACATTACACGTCCTAGC
STRF3'H1-RP7	GTGCGTCTTCAGAAACTTGG
STRF3'H2-FP2	CAAGATGAAGAAGCTTCACC
STRF3'H2-RP4	TCTCCTAAACACACAGCTTGC
STRF3'H2-FP3	ATATGTTCACAGCCGGAACG
STRF3'H2-RP3	GCGAATGTGACGAAAGTCATCC
For isolation of ORFs	
STRF3'H1-FP	ATGTCTACTTTGGGACTCACACTATTC
STRF3'H1-RP	TTAAGCCTGGTAAACTTCCTTAGCG
STRF3'H2-FP	ATGTTCTCTCCCAGCCTCG
STRF3'H2-RP	TTAAACCTGGTAAACATGGGTC
For Real time-PCR	
C.purF3'5'H-FP	ATGGCACTAGACATAGTCTTGC
C.purF3'5'H-RP	TTAAGCAACATAGGCACTTGGG
C.perF3'H1-FP3	TTGGTAGTTGGCCAAAACCG
C.perF3'H1-RP	TTAAGCCTGGTAAACTTCCTTAGCG

STRF3'H2-FPa	GAGGGAAGCTCACCGACACC
STRF3'H2-FPb	GAGGCAAGCTCACCGACACT
STRF3'H2-RPa	TTCTCCGGCAATAGCCCCTCG
STRF3'H2-RPb	TCCTCCGGCAATAGCCCGTCT
For genomic PCR	
STRF3'H1-FP4	ATGTTGACACGTGCCTTGTC
STRF3'H1-FP6	TATCCGCTATTGTGGAGGAG
STRF3'H1-RP4	TCAGTAACTGGACCATTCGC
STRF3'H1-FP3	TTGGTAGTTGGCCAAAACCG
STRF3'H2-FP1	CATCAGTGCACCTCTTTTCC
STRF3'H2-FP5	CCGTGTTGACACACAAGTTG
STRF3'H2-RP2	CTCGAGCGATATCAGCATGC
CpurF3'5'H-FP1	ACTCTGGCCGAAGGTCAGAGC
CpurF3'5'H-RP*1*	TAGTAGCCGTTCACTTCACATGC
*eEF1a*Fw	CTGGTGGTTTTGAGGCTGG
*eEF1a*Rv	CTGGCCAGGGTGGTTCATGAT

2.3 Results and discussion

2.3.1 Anthocyanins in the petals of STR

HPLC analysis of the pigments extracted from STR slip showed a particularly main peak, which was identified as Pn3Nh the major anthocyanin of STR (Figure. 2-4). However, the main anthocyanin component of *C. persicum* was identified as Mv3,5dG (Ishizaka, 2018). The precursors of these two pigments are cyanidin and delphinidin respectively. F3'H and F3'5'H are key enzymes at the branch of cyanidin and delphinidin synthetic pathway. It suggested that the red mutant of STR may be related to *F3'H* and / or *F3'5'H*.

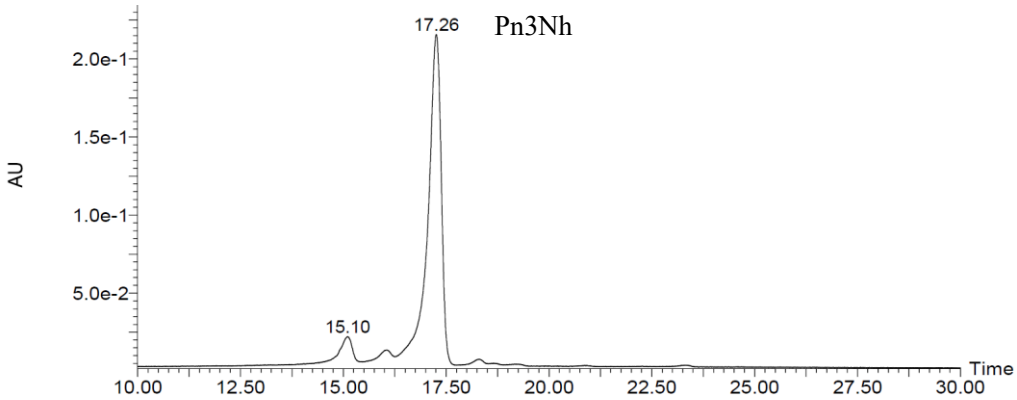


Fig. 2-4 Anthocyanins in the slips of 'Strauss', analyzed by HPLC chromatogram and detected at 530 nm

2.3.2 Isolation and sequence analysis of F3'H *genes from STR*

Two fragments of *F3'H* (*STRF3'H1* and *STRF3'H2*) were obtained from STR slips by degenerate PCR. The 5' end sequences of *STRF3'H1* was amplified by 5'RACE, the first amplification was about 700 bp, and the second amplification was about 300 bp (Figure. 2-5). The full-length cDNA of *STRF3'H1* was preliminarily obtained by splicing the two

sequences. Then gene-special primers were designed between start codon (ATG) and stop codon (TAA), to sequencing the ORF in segments, finally got the confirmed ORF of *STRF3'H1*. The 5' end sequences of *STRF3'H2* was also obtained by 5'RACE twice (Figure. 2-6). Gene-specific primers were designed based on the preliminarily obtained sequence, and the first strand of cDNA generated by reverse transcription was used as a template for amplification, finally two full-length ORFs (*STRF3'H2a* and *STRF3'H2b*) were obtained.

Gene duplication is considered a major driving of gene supplementation for secondary metabolism (Pichersky et al. 2000). Many plants contain multiple *F3'H*, like five copies of *F3'H* genes in grape genome (Castellarin et al. 2006), three copies in apple genome (Han et al. 2010). This time we also isolated three sequences from STR, the homology between *STRF3'H2a* and *STRF3'H 2b* was 99.48% (Figure. 2-7). *STRF3'H1* has 79.26%, 79.06% nucleotide sequence identity in the coding region to both *STRF3'H2a* and *STRF3'H2b* respectively. But whether these three genes have certain functions in anthocyanin synthesis needs further study. It has been reported that three *SbF3'H* genes isolated from *Sorghum bicolor*, *SbF3'H1* and *SbF3'H2* were involved in the formation of different types of anthocyanins, but *SbF3'H3* was not detected in all tissues tested (Shih et al. 2006). It suggests that although different genes were annotated as the same enzyme, these genes may represent different alternative splicing transcripts or members of gene families and do not necessarily have the same function.

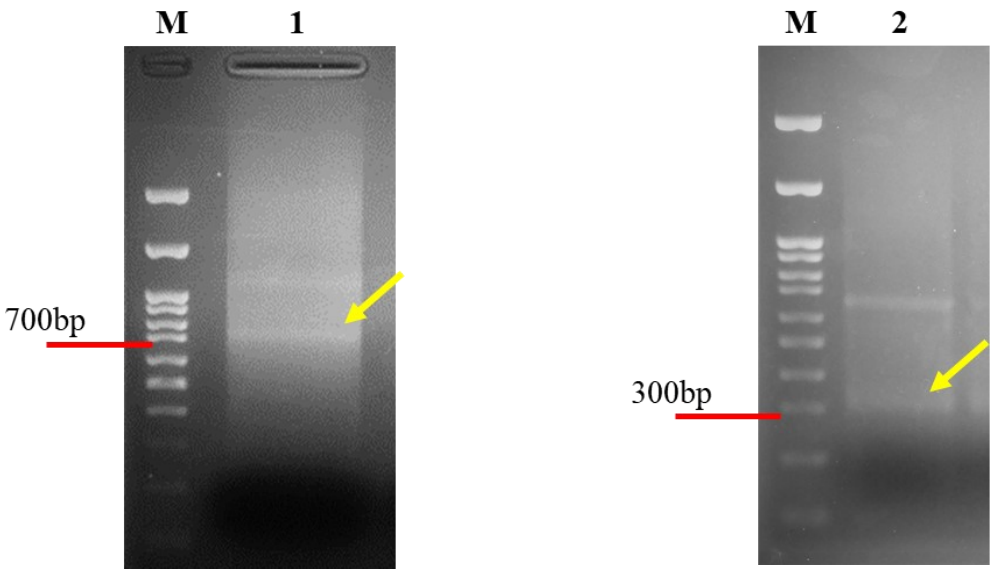


Fig. 2-5 Electropherogram of *STRF3'H1* gene 5'RACE.

M, 100bp DNA marker; 1, The first amplified fragment; 2, The second amplified fragment

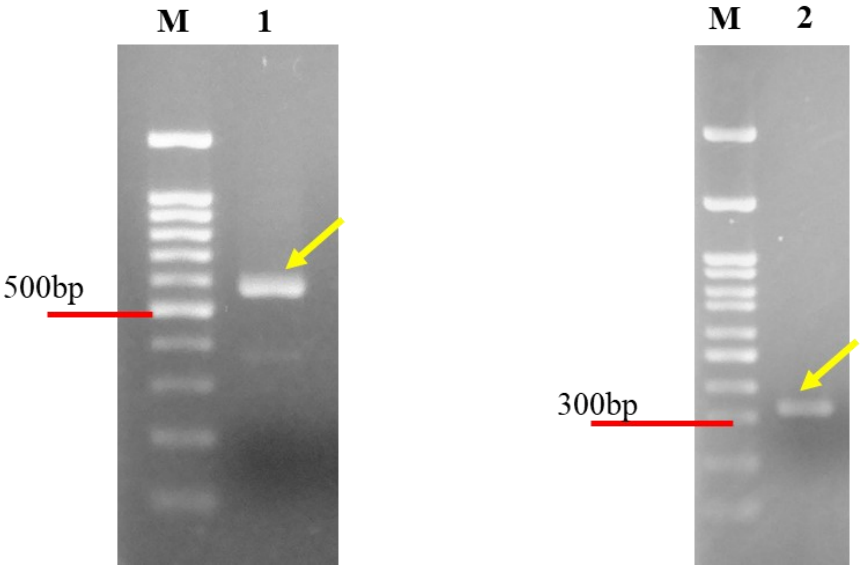


Fig. 2-6 Electropherogram of *STRF3'H2* gene 5'RACE.

M, 100bp DNA marker; 1, The first amplified fragment; 2, The second amplified fragment

```
STRF3'H2a    1 ATGTTCTCTCCCAGCCTCGCGTTCTACGCTCTAGCATTTGGGTCGTTTCTATATTATCTAATTAAATTATTATTATCTACAAAAAGTGGACAAAAGCTACGGTTACCGCCAGGGCCGAGG 120
STRF3'H2b    1 ATGTTCTCTCCCAGCCTCGCGTTCTACGCTCTAGCATTTGGGTCGTTTCTATATTATCTAATTAAATTATTATTATCTACAAAAAGTGGACAAAAGCTACGGTTACCGCCAGGGCCGAGG 120

STRF3'H2a  121 CCATGGCCGGTGATTGGGAACCTTCCACACCTCGGCCCTGTGCCACATCACTCCATCGCGGCCCTGGCACGAGTCCATGGGCCACTGATGCACCTGAGGCTCGGGGCGGTCAACGTGGTG 240
STRF3'H2b  121 CCATGGCCGGTGATTGGGAACCTTCCACACCTCGGCCCTGTGCCACATCACTCCATCGCGGCCCTGGCACGAGTCCATGGGCCACTGATGCACCTGAGGCTCGGGGCGGTCAACGTGGTG 240

STRF3'H2a  241 GTCGCTGCCTCCGCCGCTGTGGCGGCCAAGTTTTTGAAAACACACGATGCTAACTTTTCTAGCCGTCCACCAAACTCGGGCGCTAAGCATGTGGCGTATAATTATCAGGATCTTGTGTTT 360
STRF3'H2b  241 GTCGCTGCCTCCGCCGCTGTGGCGGCCAAGTTTTTGAAAACACACGATGCTAACTTTTCTAGCCGTCCACCAAACTCGGGCGCTAAGCATGTGGCGTATAATTATCAGGATCTTGTGTTT 360

STRF3'H2a  361 GCACCGTACGGACCGAGGTGGCGGTTGCTGAGGAAGATATCATCAGTGCACCTCTTTTCGGGCAAGGCGTTGGATGACTTTCGTCACATTCGCGAGGAAGAGGTAGCAGTGTTGACACAC 480
STRF3'H2b  361 GCACCGTACGGACCGAGGTGGCGGTTGCTGAGGAAGATATCATCAGTGCACCTCTTTTCGGGCAAGGCGTTGGATGACTTTCGTCACATTCGCGAGGAAGAGGTAGCAGTGTTGACACAC 480

STRF3'H2a  481 AAGTTGGTGGGTCGGGCTCAGTCGCAAGCTGTGTGTTTAGGAGAACTGCTTAACGTGTGTACTGCCAACGCGCTCTCGCGTGTCATGCTAGGGCGGCGTGTTCTCAGTGATGCAAGCGAA 600
STRF3'H2b  481 AAGTTGGTGGGTCGGGCTCAGTCGCAAGCTGTGTGTTTAGGAGAACTGCTTAACGTGTGTACTGCCAACGCGCTCTCGCGTGTCATGCTAGGGCGGCGTGTTCTCAGTGATGCAAGCGAA 600

STRF3'H2a  601 AACAAGGGCGATTTGAAGGCGAACGAGTTCAAGGAGATGGTGGTGGAGCTGATGGTGCTCGCTGGAGTTTTCAACATCGGAGATTTCATTCCGTCGTTAGAGTGGCTCGACCTACAGGGT 720
STRF3'H2b  601 AACAAGGGCGATTTGAAGGCGAACGAGTTCAAGGAGATGGTGGTGGAGCTGATGGTGCTCGCTGGAGTTTTCAACATCGGAGATTTCATTCCGTCGTTAGAGTGGCTCGACCTACAGGGT 720

STRF3'H2a  721 GTTGCTGCCAAGATGAAGAAGCTTCACCGTCGCTTTGATGCCTTCCTCGCAGCCATTGTCGAGGAGCACAAGGCCGAGGCCGCCTTGGGCGATGATAAAAAACAAAGGGACTTTCTTAGC 840
STRF3'H2b  721 GTTGCTGCCAAGATGAAGAAGCTTCACCGTCGCTTTGATGCCTTCCTCGCAGCCATTGTCGAGGAGCACAAGGCCGAGGCCGCCTTGGGCGATGATAAAAAACAAAGGGACTTTCTTAGC 840

STRF3'H2a  841 ATGCTGATATCGCTCGAGGATGATGGCGAGGGAGGGAAGCTCACCGACACGGAGATCAAAGCCCTGCTCTTGAATATGTTCACAGCCGGAACGGACACTTCATCAAGTACCGTCGAATGG 960
STRF3'H2b  841 ATGCTGATATCGCTCGAGGATGATGGCGAGGGAGGCAAGCTCACCGACACTGAGATCAAAGCCTTGCTCTTGAATATGTTCACAGCCGGAACGGACACTTCATCAAGTACCGTCGAATGG 960

STRF3'H2a  961 GCCATAGCTGAACTCATCCGTCACCCGAAAATTCTAGCCCAAGCTCAAATAGAATTGGACTCGGTAGTCGGACCGAACCGCCTTGTGTCCGACTCGGACTTGTCCCAACTACCCTTCCTC 1080
STRF3'H2b  961 GCCATAGCTGAACTCATCCGTCACCCGAAAATTCTAGCCCAAGCTCAAATAGAATTGGACTCGGTAGTCGGACCGAACCGTCTTGTGTCCGACTCGGACTTGTCCCAACTACCCTTCCTC 1080

STRF3'H2a 1081 CAAGCTGTCATCAAGGAGACCTTCAGGCTACACCCCTCAACTCCCCTGTCTCTTCCTCGTATGGCCTCCGAGAGCTGCGAGATCGACGGATACCACATCCCCCAGGGCTCCACTCTCCTT 1200
STRF3'H2b 1081 CAAGCTGTCATCAAGGAGACCTTCAGGCTACACCCCTCAACTCCCCTGTCTCTTCCTCGTATGGCCTCCGAGAGCTGCGAGATCGACGGATACCACATCCCCCAGGGCTCCACTCTCCTT 1200

STRF3'H2a 1201 GTGAATGTCTGGGCCATAGCCCGCGACCCGGATGCATGGGCCGAGCCCCTAGAGTTTAGGCCCGAGCGGTTCTTGCCAGGTGGAGAGAAGCCCAATGCCGATGTCAGAGGGAATGATTTT 1320
STRF3'H2b 1201 GTGAATGTCTGGGCCATAGCCCGCGACCCGGATGCATGGGCCGAGCCCCTAGAGTTTAGGCCCGAGCGGTTCTTGCCAGGTGGAGAGAAGCCCAATGCCGATGTCAGAGGGAATGATTTT 1320

STRF3'H2a 1321 GAGGTTATACCATTTGGGGCTGGTCGCAGAATATGTGCCGGGATGAGCCTCGGGTTGCGTATGGTCCAGCTACTGACAGCGACACTGGTCCATGGGTTCGACTGGGTCCTGGCCGAGGGG 1440
STRF3'H2b 1321 GAGGTTATACCATTTGGGGCTGGTCGCAGAATATGTGCCGGGATGAGCCTCGGGTTGCGTATGGTCCAGCTACTGACAGCGACACTGGTCCATGGGTTCGACTGGGTCCTGGCAGACGGG 1440

STRF3'H2a 1441 CTATTGCCGGAGAAGTTGAATATGGACGAGGCGTATGGTTTGACCTTGCAACGAGCTGTGCCACTTATGGTCCACCCCCAACCGAGGCTACCGACCCATGTTTACCAGGTTTAA      1554
STRF3'H2b 1441 CTATTGCCGGAGGAGTTGAATATGGACGAGGCGTATGGTTTGACCTTGCAACGAGCTGTGCCACTTATGGTCCACCCCCAACCGAGGCTACCGACCCATGTTTACCAGGTTTAA      1554
```

Fig. 2-7 ORFs of *STRF3'H2a* and *STRF3'H2b*. The difference between these two genes is only 8 bases in the downstream.

2.3.3 Structural features and homology analysis of STRF3'H protein

Three candidate genes deduced to encode 507, 517, 517 amino acids separately. The predicted molecular weights of these three amino acids were 56.1, 56.9, 56.9 kDa, and the calculated isoelectric point were 6.82 ,7.03, 6.66, respectively. The instability index (II) of F3'H1 was computed to be 34.49, which can be classified as a stable protein. On the contrary, F3'H2a and F3'H2b were classified as unstable protein because the instability index (II) was computed to be 41.26 and 40.97. The grand average of hydropathicity (GRAVY) of F3'H1 was computed to be -0.010, which is a hydrophilic protein. However, the GRAVY of both F3'H2a and F3'H2b was computed to be 0.031, which means that they are hydrophobic proteins.

SOPMA predicted that the secondary structure of STRF3'H1 is mainly made up of alpha helix (46.75%) and random coils (37.08%), while extended strands (10.45%) and beta turn (5.72%) only occupied a small part. STRF3'H2a and STRF3'H2b also have similar structural characteristics. STRF3'H2a is composed of alpha helix (46.81%), random coils (35.98%), extended strands (11.80%) and beta turn (5.42%). STRF3'H2b is composed of alpha helix (47.39%), random coils (36.56%), extended strands (11.61%) and beta turn (4.45%). TMHMM online analysis of the amino acid sequences showed that STRF3'H2a and STRF3'H2b have parallel structures, both of them have a transmembrane domain located at S_5 to T_{27}, the part of M_1 to P_4 located inside the microsomal membrane, K_{28} to V_{517} located outside (Figure. 2-8a, b). However, the prediction also showed that STRF3'H1 has no obvious transmembrane, and the entire peptide chain may be located outside membrane (Figure. 2-8c). Arabidopsis TT7 was also analyzed, the results were same as STRF3'H2a and STRF3'H2b (Figure. 2-8d). Previous research has shown that microsomal-type cytochrome P450 is a type of membrane protein that bounds to the membrane through

the N-terminal transmembrane signal anchor domain to exert physiological functions (Murakami et al. 1994). From amino acid sequence alignment, the N-terminal of F3'H1 was five amino acid residues less than that of F3'H2a and F3'H2b. Does this affect the anchoring of F3'H1 on membrane thus the structure and function? Actually, the functional analysis based only on the amino acid sequence is not very rigorous, and functional verification of F3'H is still needed.

The amino acid sequence alignment of STRF3'H1, STRF3'H2a, STRF3'H2b with TT7 from Arabidopsis showed 68.2%, 66.4% and 66.2% identities; with PhF3'H (*Petunia hybrida*) showed 70.8%, 70.3% and 70.1% identities; with VaF3'H (*Vitis amurensis*) showed 75.5%, 75.4% and 74.8% identities (Figure. 2-9). And it's obvious that all three deduced protein sequences contain several domains highly conserved in plant F3'Hs. Three F3'H-specific conserved motifs were found in STRF3'Hs amino acid sequence. All of three STRF3'H amino acid sequences have the motif that exactly same as "VVVAAS" (Brugliera et al. 1999), but "GGEK" was present at $G_{417}GER_{420}$ of *STRF3'H1*, "VDVKG" were present at "$A_{423}DVRG_{427}$" of *STRF3'H1*, "$A_{433}DVRG_{437}$" of *STRF3'H2a* and *STRF3'H2b*. In addition, four cytochrome P450-specific conserved motifs were also found. The heme domain "FGAGRRICAG", which is considered to be responsible for the enzyme to bind carbon monoxide (Werck-Reichhart et al. 2002), was present in all three sequences. STRF3'Hs also bear the conserved motif "AGTDTS", forming an oxygen-binding pocket, which is required for its catalytic activity (Chapple 1998). There is an E_{356}-R_{359}-R_{398} (*STRF3'H1*) trinity forms a pocket lock motif to stabilize the core structure, while it also presents at *STRF3'H2a*, *STRF3'H2b* "E_{366}-R_{369}-R_{408}" (Hasemann et al. 1995). The proline-rich region "PPGPNPWP" considered to be the hinge motif necessary for P450 enzyme to anchor on the membrane, it was existed in the three genes (Murakami et al. 1994; Yamazaki et al. 1993). However, TMHMM speculated that F3'H2a and F3'H2b each have a

transmembrane domain located at the 5th to the 27th amino acid, while F3'H1 may not. According to amino acid sequence alignment, the N-terminal of F3'H1 was 5 amino acid residues less than that of F3'H2a and F3'H2b. Whether this affects the anchoring of F3'H1 on the membrane and thus affects its structure and function needs to be further explored. "GGEK" is a unique motif to F3'H, an important feature that distinguishes them from their close relatives F3'5'H (Brugliera et al. 1999). But in sorghum and other monocots (except rice), it mostly presented as "GGSH" (Boddu et al., 2004). In this study the F3'H unique motif in STRF3'H1 presented as "GGER", which were also found in *Vitis vinifera* (Castellarin et al. 2006), *Antirrhinum kelloggii* (AB547161) and other plants.

A phylogenetic tree of several reported plant F3'H and F3'5'H protein sequences was shown in figure 2-10, STRF3'H1, STRF3'H2a, STRF3'H2b were clustered in the CYP75B clade together with other F3'Hs but further from Cyclamen F3'5'H group in evolutionary distance. The phylogenetic tree and multiple alignment strongly suggested that *STRF3'H1*, *STRF3'H2a*, *STRF3'H2b* encodes a typical flavonoid 3'-hydroxylase.

2.3.4 Genomic structure analysis of STRF3'Hs

To analyze the genomic structure of *STRF3'H* genes in detail, the genomic *STRF3'H* genes were separated into four regions (A, B, C, D) and amplified with four pairs of gene-specific primers. When pairwise aligning with the full-length cDNA sequence, one intron was found in *STRF3'H1* (Figure. 2-11a), two introns were found in *STRF3'H2a and STRF3'H2b* (Figure. 2-12a), respectively. All the introns follow standard splicing sites, starting at GT and ending at AG. *STRF3'H2a* and *STRF3'H2b* both have two introns, they have same structure and only showed 8 bp differs in the 3rd exon. While *STRF3'H1* has only one intron, the position almost same to the first intron of *STRF3'H2a* and *STRF3'H2b*. Previous studies also revealed that the length of the intron of *F3'Hs* varies greatly, but the

position is relatively conservative. Most of the first intron is located 400-500 bp after the start codon. The number of introns of the plants *F3'H* gene is not fixed neither. It is common to have two introns, such as the *F3'Hs* of grape (Jeong et al. 2006), morning glory (Hoshino et al. 2003), apple (Han et al. 2010), while the *Poaceae* usually have one intron, like *F3'H* of sorghum (Boddu et al., 2004). According to a recent report, *GbF3'H1* cloned from *Ginkgo biloba* L. also contains only one intron (Wu et al., 2020). Many plants have multiple introns in the *F3'H* gene, for example three introns in *Arabidopsis thaliana TT7* (AF155171) and three in *Brassica napus F3'H* (Xu et al. 2007). In a rare case, *IbF3'H* gene (*Ipomoea batatas* L. Lam) has no introns (Zhou et al. 2012).

Differences in *STRF3'Hs* between 'Strauss' and *C. persicum* were investigated at the genomic level. The same four sets of *STRF3'H1*-specific primers were used to amplify the entire *F3'H1* of *C. persicum*, and four DNA fragments of the same size as those amplified from 'Strauss' were obtained (Figure. 2-11b). The analysis of *STRF3'H2a, 2b* also presented the similar results (Figure. 2-12b).

2.3.5 Genomic PCR of CperF3'5'H

Based on the cDNA sequence of *CperF3'5'H* (GQ891056), two pairs of *CperF3'5'H*-specific primers were designed to amplify the entire gene of *C. persicum*. There was only one intron in *CperF3'5'H* (Figure. 2-13a). When we used the same two sets of *CperF3'5'H*-specific primers to amplify the entire *F3'5'H* of 'Strauss', two DNA fragments of the same size as those amplified from *CperF3'5'H* were obtained (Figure. 2-13b). These indicated that 'Strauss' and *C. persicum* were not different at the genomic level. However, it is worth noting that we only compared the genomic sequences within ORF, and whether there are differences in the cis-acting elements, still need to be further explored.

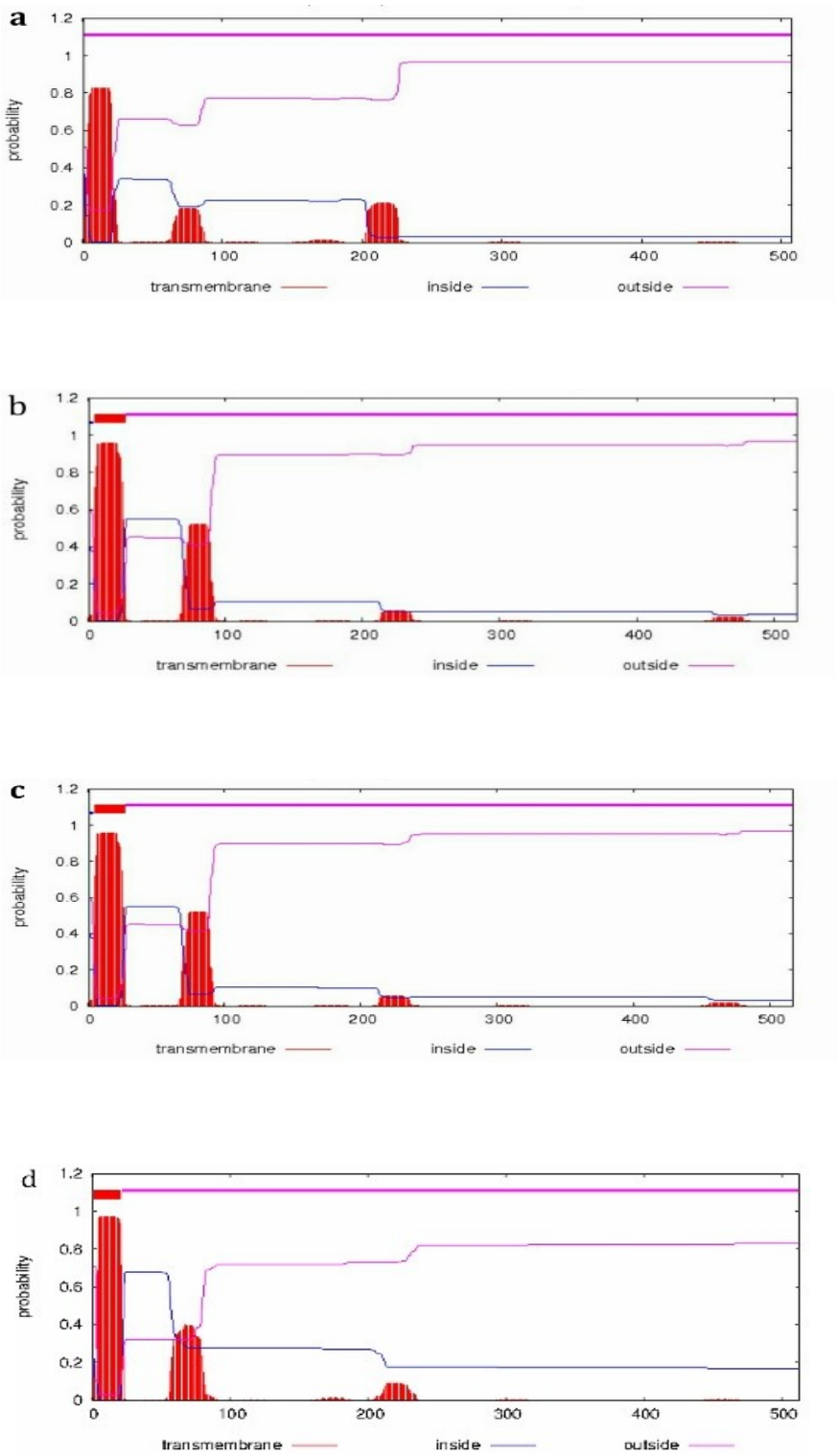


Fig. 2-8 Predicted transmembrane structure domain of *STRF3'H1*(a), *STRF3'H2a* (b), *STRF3'H2b*(c), *tt7(d)*

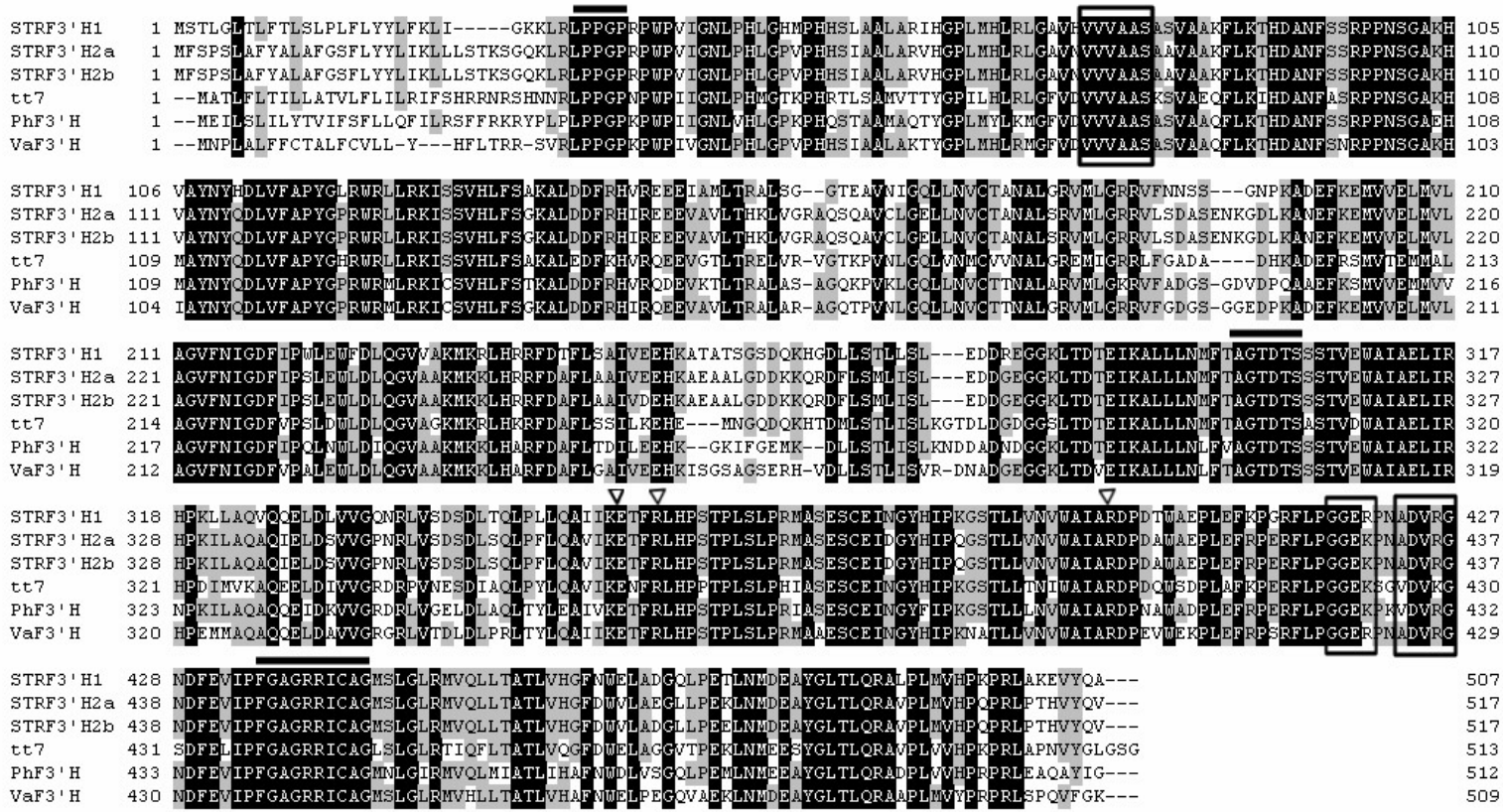


Fig. 2-9 Multiple alignment of the deduced amino acid sequences of *F3'H1*, *F3'H2a*, *F3'H2b* from STR and *Arabidopsis TT7* (AF155171), *PhF3'H* (AF155332), *VaF3'H* (FJ645766). The F3'H-specific motifs "VVVAAS", "GGEK", and "VDVKG" are marked with black boxes. The black lines indicate cytochrome P450-specific conserved motifs. Arrowheads indicate an E-R-R triad forming the pocket locking motif for the stabilization of the core structure.

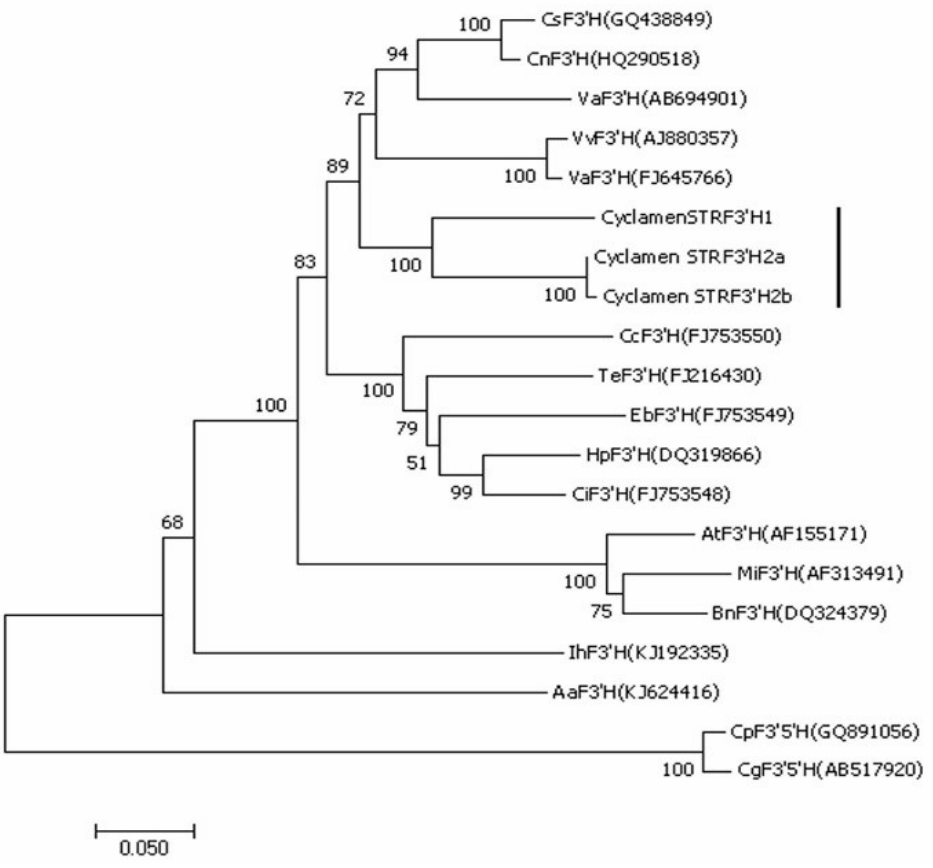


Fig. 2-10 Phylogenetic tree analysis of STRF3'H and other known F3'Hs and F3'5'Hs. Vertical line indicates *F3'Hs* isolated from STR in this study. *CsF3'H*, *Camellia sinensis*; *CnF3'H*, *Camellia nitidissima*; *VaF3'H, Vaccinium ashe*; *VvF3'H*, *Vitis vinifera*; *CcF3'H*, *Centaurea cyanus*; *TeF3'H*, *Tagetes erecta*; *EbF3'H*, *Echinops bannaticus*; *HpF3'H*, *Hieracium pilosella*; *CiF3'H*, *Cichorium intybus*; *AtF3'H(TT7)*, *Arabidopsis thaliana*; *MiF3'H*, *Matthiola incana*; *BnF3'H*, *Brassica napus*; *IhF3'H*, *Iris x hollandica*; *AaF3'H*, *Anthurium andraeanum*; *CpF3'5'H*, *Cyclamen persicum*; *CgF3'5'H*, *Cyclamen graecum.*

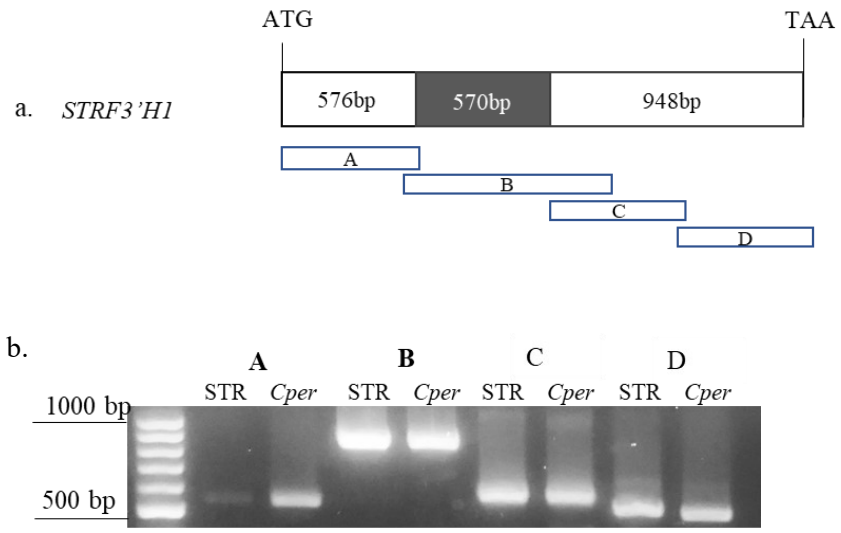


Fig. 2-11 Genomic analysis of *STRF3'H1*. (a) The genomic structures of *STRF3'H1*. The open boxes indicate exons, the shadowed box indicate introns. Boxes A, B, C and D indicate the amplified regions that separate *STRF3'H1* into four parts. (b) Genomic PCR of four separated regions of *STRF3'H1* in 'Strauss' and *C. persicum*. A, B, C and D are as shown in (a).

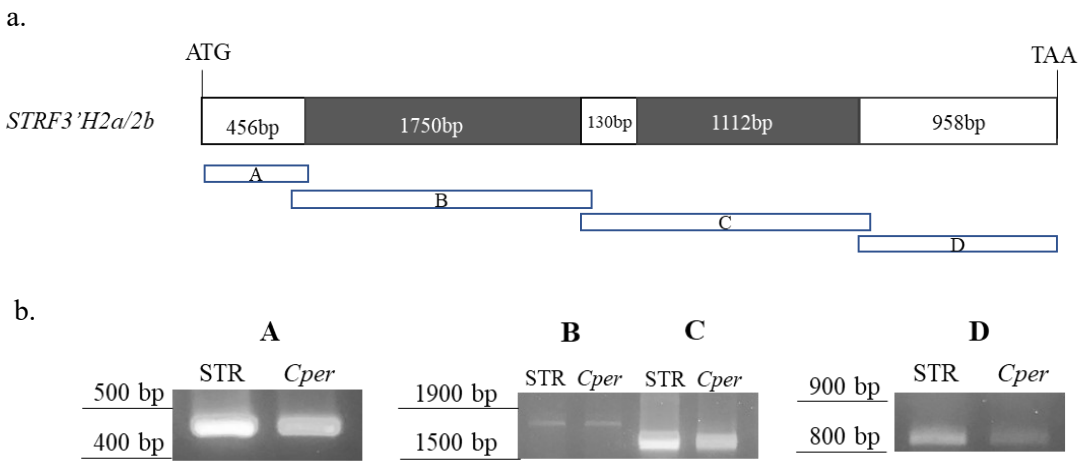


Fig. 2-12 Genomic analysis of *STRF3'H2a* and *STRF3'H2b*. (a) The genomic structures of *STRF3'H2a* and *STRF3'H2b*. The open boxes indicate exons, the shadowed boxes indicate introns. There are 8 bp differs in their 3rd exon. Boxes A, B, C and D indicate the amplified regions that separate *STRF3'H2a/b* into four parts. (b) Genomic PCR of four separated regions of *STRF3'H2a/b* in 'Strauss' and *C. persicum*. A, B, C and D are as shown in (a).

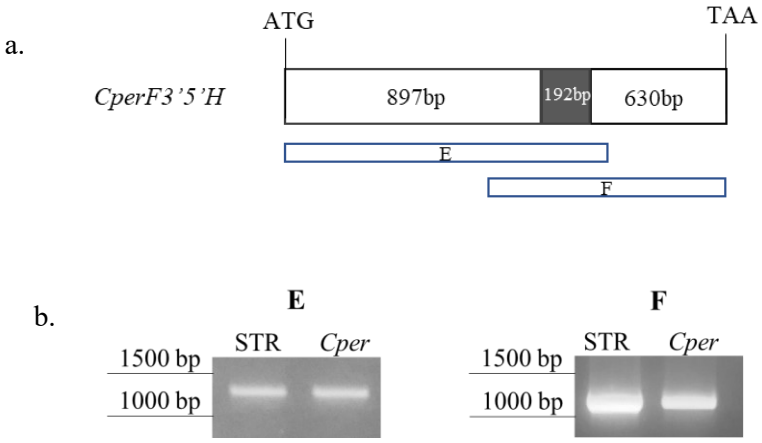


Fig. 2-13 Genomic analysis of *CperF3'5'H*. (a) The genomic structure of *CperF3'5'H*. The open boxes indicate exons, the shadowed box indicate intron. Boxes E and F indicate the amplified regions that separate *CperF3'5'H* into two parts. (b) Genomic PCR of two separated regions of *CperF3'5'H* in '*Strauss*' and *C. persicum*. E and F are as shown in (a).

2.3.5 Expression pattern of STRF3'Hs

The expression profiles of *STRF3'Hs* at different developmental stages of STR petals were investigated. As the results shown *STRF3'Hs* expressed stronger in the early stage and weakened gradually with the coloring of petals, which was same to *Petunia* (Brugliera et al. 1999), *Eustoma grandiflorum* (Takatori et al. 2005) (Figure. 2-14). Considering the expression patterns, genes encoding enzymes involved in flavonoid biosynthesis in *Gentiana* were classified into three groups: *CHS* and *CHI*, common enzymes necessary for biosynthesis of both flavone and anthocyanin are belong to Group I, they are expressed in all stages of flower development; Group II included *F3'H* and *FSII*, which expressed in early stages of flower development; Group III included *F3H*, *F3'5'H*, *DFR*, *ANS*, *3GT* and *5AT*, which are expressed in late stages of flower development (Nakatsuka et al. 2005). *STRF3'Hs* in this study present same expression patter as that *GtF3'H*. However, *F3'H* from azalea showed a different trend, it expressed weakly at the first stage, but got the highest level at the second stage, and then decreased. The author considered that the expression of *F3'H* is not directly related to the accumulation of anthocyanins, because *F3'H* also abundantly expressed in some non-pigment organs (Nakatsuka et al. 2008). The transcription level of *F3'H* in the petals of blue torenia (SWB) was higher in the second and third stages, and these two stages were the highest accumulation levels of flavones and anthocyanins, while the transcription levels in other stages were almost not detected. However, the content of cyanidin-type-anthocyanidins was higher in transgenic torenia (obtained by over-expression of the *F3'H* in SWB), the expression of *F3'H* was detected at all stages of flowering, and the expression of *F3'H* gradually decreased with the opening of flowers (Ueyama et al. 2002). It seems that the expression of *F3'H* is closely related to the content of flavones and anthocyanins, especially cyanidin-type-anthocyanidins.

The expression of *STRF3'H* genes was detected in the leaves. *STRF3'H2a* in leaves expressed weaker than that at stage1 and stage2, but stronger than the full-opened stage. *STRF3'H2b* in leaves expressed weaker than stage1, but higher than the other three stages. The expression level of *STRF3'H1* was higher than that of any flowering stage. High expression of *F3'H* has also been detected in the leaves of other plants, such as *Arabidopsis* (Schoenbohm et al. 2000), *Perilla* (Kitada et al. 2001) and *Azalea* (Nakatsuka et al. 2008). The expression levels of *GbF3'H1* in ginkgo leaves were significantly higher than that in flowers, roots and other tested tissues, which because of this expression is associated with the secondary metabolism product accumulation in the corresponding organs (Wu et al., 2020). Although we know that the expression of *F3'H* is related to the accumulation of anthocyanins, there is almost no accumulation of anthocyanins in leaves, which may be caused by insufficient expression of other genes in the anthocyanin synthesis pathway (Noda et al. 2004). On the other hand, this also suggests that *F3'H* plays a role in the synthesis of other secondary metabolites.

2.3.6 Expression analysis of F3'Hs *and* F3'5'H *in* C. persicum *and STR*

The transcription level of *F3'Hs* genes and *F3'5'H* gene in *C. persicum* and STR were analyzed by real-time PCR. *STRF3'5'H* was cloned from *C. persicum* (GQ891056). Gene special primer pairs were applied to compare the expression levels of these genes in two materials. Expression levels of *F3'Hs* and *F3'5'H* transcripts were relative to those of *eEF1α*. As shown in figure 2-15, the expression level of *F3'Hs* in STR were higher than those in *C. persicum*, especially the relative expression of *F3'H1* in STR is 25 times that of *C. persicum*. *F3'5'H* was expressed strongly in *C. persicum* about 33 times that of STR.

Variation in flower color is generally the result of differences in either structural or regulatory genes involved in the flavonoid biosynthetic pathway (Tanaka et al. 2008). Compared with *C. persicum*, STR has a completely different main anthocyanin composition, so we compared the expression level of key genes related to their main anthocyanin synthesis. The expression level of *F3'5'H* in *C. persicum* was higher than that in STR, and the main anthocyanin component of *C. persicum* was Mv3,5dG. Previous study revealed that F3'5'H is responsible for synthesizing delphinidin and its derivatives, *F3'5'H* gene encoding this enzyme often referred to as the blue gene (Holton et al. 1994). Some plants, such as chrysanthemums, roses and carnations etc., cannot produce blue or purple flowers, partly because they lack the activity of the F3'5'H enzyme (Tanaka and Brugliera 2013). Loss of endogenous *F3'5'H* transcript will cause the flower color of cyclamen to change from purple to red/pink (Boase et al. 2010). In the present study, is the reduced expression level of *F3'5'H* of STR responsible for the lack of Mv3,5dG accumulation in its flowers? There are two main causes of defective gene expression, one is a mutation in the gene itself and the other is a mutation in the transcription factor that regulates the expression of that gene. For example, due to the insertion of a retrotransposon in the first exon, *RsF3'H* (*Raphanus sativus* L.) lost its function and promotes the accumulation of pelargonidin-bases anthocyanin (Masukawa et al. 2018). We investigated the differences in *CperF3'5'H* between *C. persicum* and STR at the genomic level, and found no difference. Therefore, whether the apparently reduced expression of *F3'5'H* in STR is caused by the mutation of the associated transcription factors needs to be further explored.

Having said that, the red petals of STR contain a large number of Pn3Nh, and *F3'Hs* was expressed extensively in STR. *F3'H* is a vital structural gene of flavonoid anabolism, which competes with FLS and DFR for substrates and controls the synthesis direction of

flavonoids towards the synthesis of cyanidin-based anthocyanins (Olsen et al. 2010). The *Arabidopsis transparent testa 7* (*tt7*) mutant that has no F3'H function exhibits a yellow seed coat, and the anthocyanin accumulation level of the mutant plant is lower compared to wild species (Schoenbohm et al. 2000). Castellarin et al. (2006) found that the expression of *F3'H* was associated with the accumulation of cyanidin-based anthocyanins in the berry skin of grapevines. Overexpression of *F3'H* gene in *I. nil* cultivar 'Violet' can change the flower color of 'Violet' that lacks F3'H function from red to blue (Takatori et al. 2015). A recent study found that an indel in *PsF3'H* (*Paeonia suffruticosa* Andr.) affects the synthesis of cyanidin-based anthocyanins, and this may be the reason of the low anthocyanin content in acyanic petals. *RcMF3'H1* was related to synthetic cyanidin-based red anthocyanins (Zhang et al. 2020). In this study the high expression level of *F3'Hs* and large accumulation of red pigment in STR slips suggested that F3'H may play a pushing role in the flower color formation, although this view still needs to be determined. However, the expression levels of these *STRF3'H* genes were also different, further research is needed to determine which one is a functional.

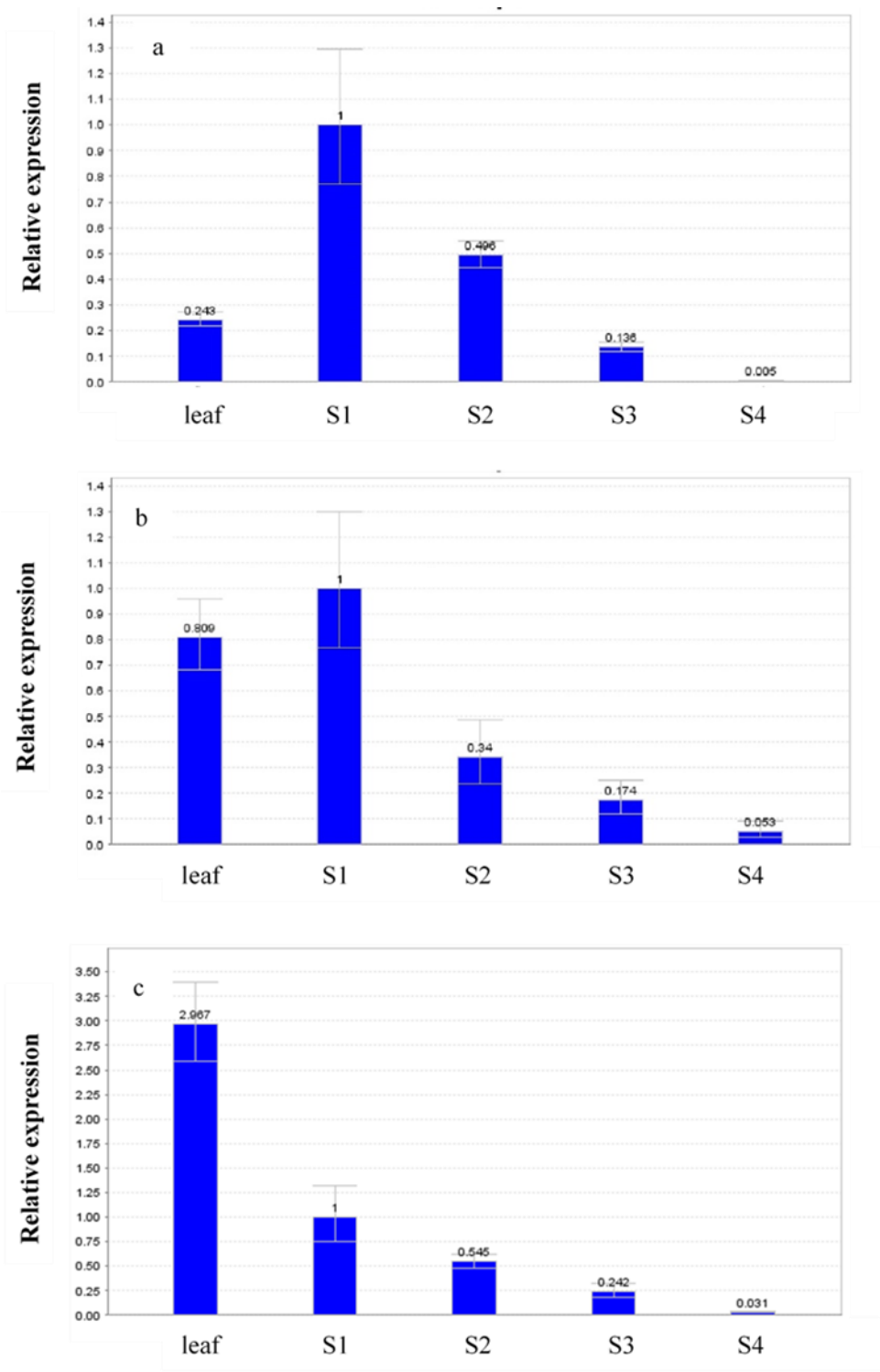


Fig. 2-14 Relative expressions of *STRF3'H* genes in leaves and five flowering stages of 'Strauss'. (a) *STRF3'H1* (b) *STRF3'H2a* (c) *STRF3'H2b. eEF1α* gene was amplified as an internal control. The error bar represents the standard error. S1,S2,S3,S4: developmental stages of STR petals.

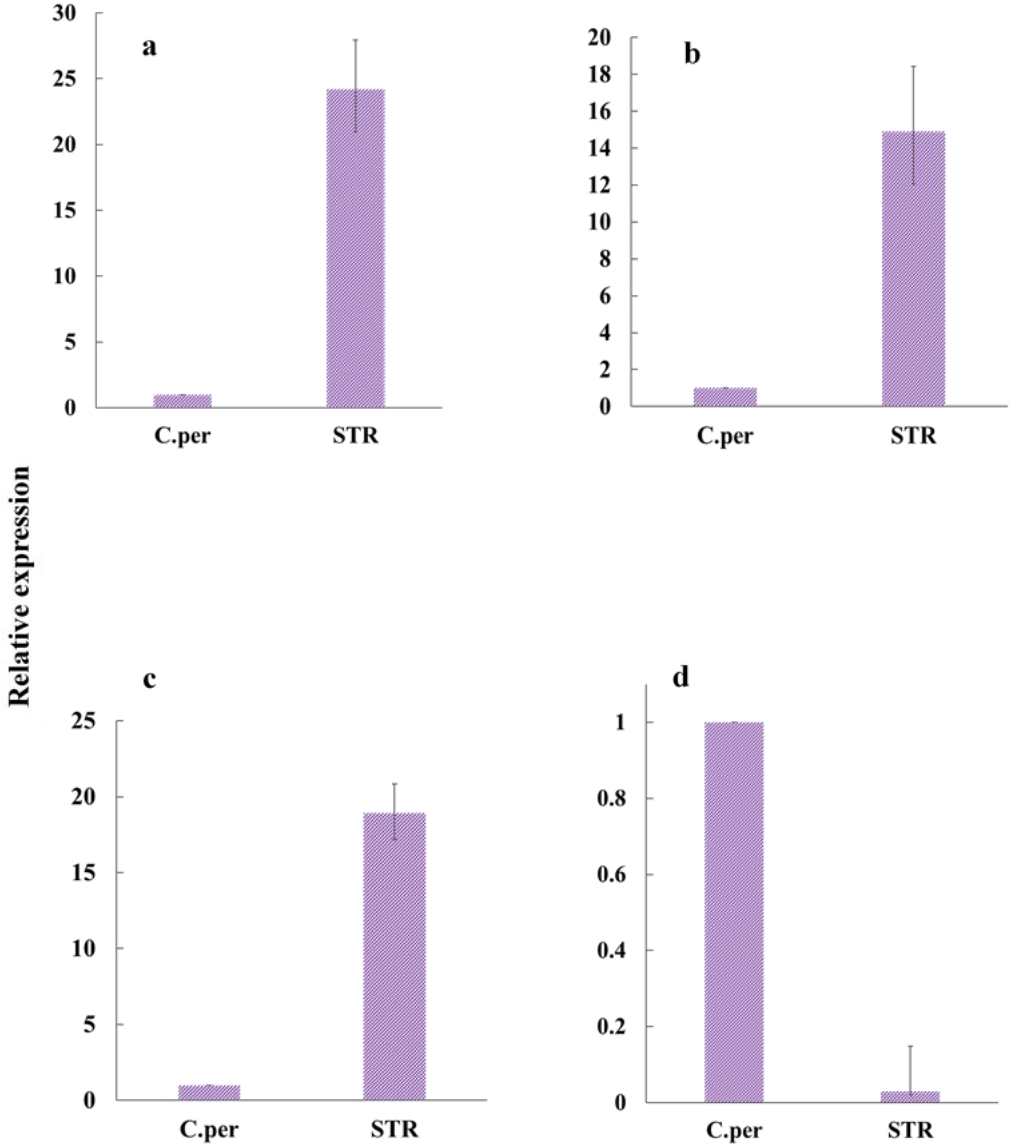


Fig. 2-15 Expression analysis of *F3'H* genes, *F3'5'H* gene in *C. persicum* and STR.

(a) *STRF3'H1* (b) *STRF3'H2a* (c) *STRF3'H2b* (d) *F3'5'H*

eEF1α gene was amplified as an internal control. The error bar represents the standard error.

C.per, *C.persicum*; STR, 'Strauss'

2.3.7 In vitro expression of STRF3'Hs

The recombinant vectors (pET21a-*STRF3'H1*, pET21a-*STRF3'H2a*, pET21a-*STRF3'H2b*), were constructed and sequenced, and the sequences of the inserted genes were determined to be accurate. Transformed the plasmids with correct sequence into *E. coli* BL21(DE3) by heat shock, prepare for protein induction expression.

By comparing different induction temperature, time, and IPTG concentration, it was finally decided that 20 °C, 16 hours, 0.1 mM was the best induction condition for pET21a-*STRF3'H1*; 20 °C, 16 hours, 0.5 mM was the best induction condition for pET21a-*STRF3'H2a* and pET21a-*STRF3'H2b*. After induction according to the optimal conditions, the cells were collected and lysed, the supernatant was recovered for SDS-PAGE. The supernatant of the crude enzyme lysate of pET21a-*STRF3'H2a* and the purified supernatant were subjected to electrophoresis. As a result, a significant band was detected between 66kDa and 45kDa compared to the control without IPTG induction (Figure 2-16). Similar results were obtained by electrophoresis of crude enzyme lysate supernatant and purified supernatant of pET21a-*STRF3'H2b* (Figure 2-17). These indicated that the STRF3'H2a and STRF3'H2b protein can be expressed normally in *E. coli* under IPTG induction. However, SDS-PAGE of the supernatant of pET21a-*STRF3 H1* crude enzyme lysate showed almost no bands, and its precipitation electrophoresis showed clear target bands with similar size to the predicted protein (Figure 2-18). It seemed that the recombinant protein of pET21a-*STRF3'H1* mostly exists in the form of inclusion bodies. There are many reasons for the insoluble expression of proteins, such as pressure from the external environment and biological limitations. Next, we will continue the attempts to reduce misreading in translation and synthesis by choosing proper composition, concentration and proportion of culture medium and proper temperature.

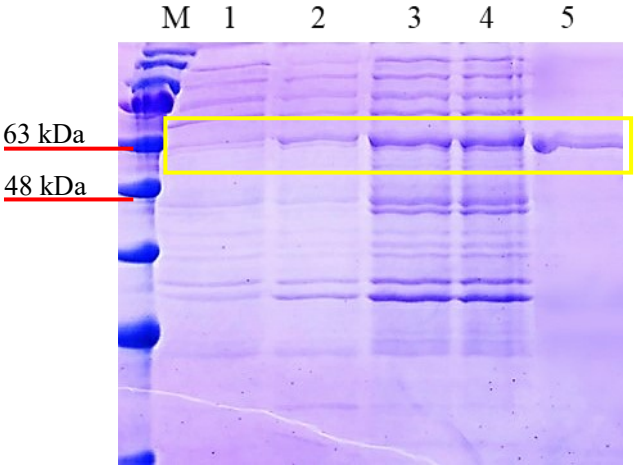


Fig. 2-16 SDS-PAGE analysis of the recombinant pET21a-*STRF3'H2a*

Lane M: molecular mass marker; lane1, supernatant after 0 M IPTG induction; lane 2, supernatant after 0.1 mM IPTG induction; lane 3, supernatant after 0.5 mM IPTG induction; lane 4, supernatant after 1 mM IPTG induction; lane 5, purified supernatant after 0.5 mM IPTG induction. Induction condition: 20 ℃, 16 h.

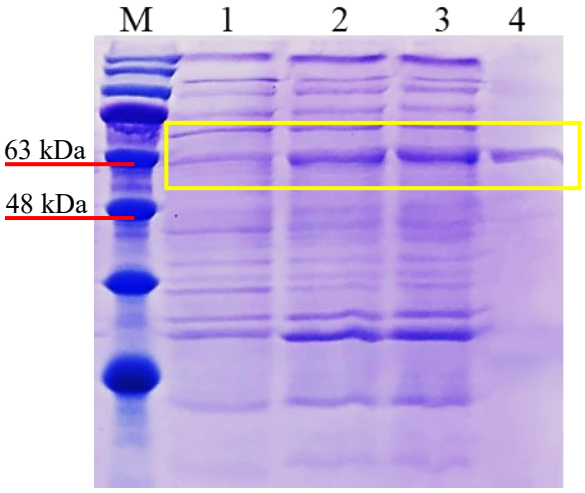


Fig. 2-17 SDS-PAGE analysis of the recombinant pET21a-*STRF3'H2b*

Lane M: molecular mass marker; lane1, supernatant after 0 M IPTG induction; lane 2, supernatant after 0.1mM IPTG induction; lane 3, supernatant after 0.5 mM IPTG induction; lane 4, purified supernatant after 0.5 mM IPTG induction.

Induction condition: 20 ℃, 16 h.

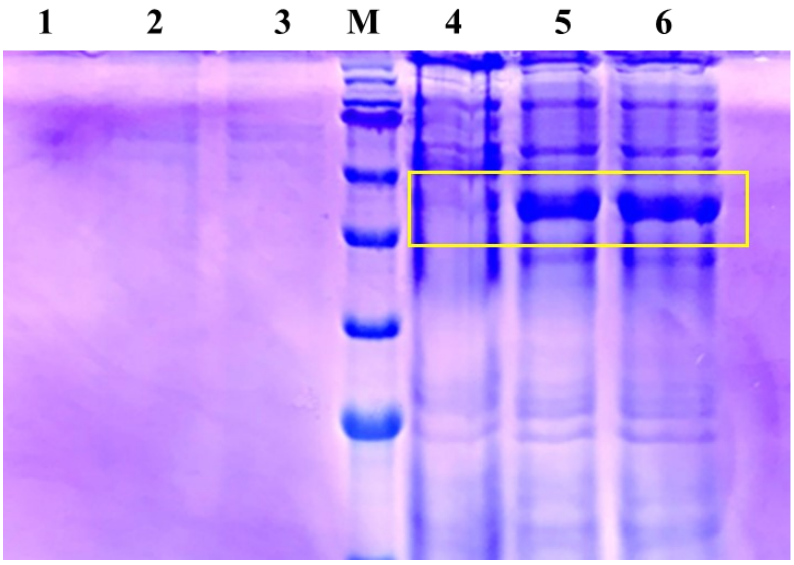


Fig. 2-18 SDS-PAGE analysis of the recombinant pET21a-*STRF3'H1*

Lane M: molecular mass marker; lane1, supernatant after 0 M IPTG induction; lane 2, supernatant after 0.1mM IPTG induction; lane 3, supernatant after 0.5 mM IPTG induction; lane 4, precipitation after 0 M IPTG induction; lane 5, precipitation after 0.1mM IPTG induction; lane 6, precipitation after 0.5mM IPTG induction.

Induction condition: 20 ℃, 16 h.

2.4 Concluding remarks

F3'H and F3'5'H are the key enzymes involved in anthocyanin biosynthesis, which affect the direction of pigment formation and the type of final pigment. In this study, three ORFs of *F3'H* gene were cloned from STR by RACE method. The results of real-time PCR showed that the expression level of STR*F3'H*s were the highest at the early stage of flower development, decreased with flower development and reached the lowest at the flowering stage. Therefore, we speculate that STRF3'Hs are involved in the formation and accumulation of anthocyanin precursors. *STRF3'Hs* also expressed strongly in leaves with very low anthocyanin content, which indicates that *F3'H* also plays an important role in other secondary metabolism pathways besides anthocyanin.

STR is a cultivar of *C. persicum*, the mainly pigment of STR has been changed from Mv3,5dG to Pn3Nh. F3'H and F3'5'H are the key enzymes in these two branches of pigment synthesis, so we compared the expression of *F3'H* and *F3'5'H* in these two flowers. The expression level of *F3'H* in red STR is much higher than *F3'5'H*, while *F3'5'H* is expressed stronger in purple *C. persicum* and weaker in STR. This suggested that the variation of STR flower color is related to F3'H and/or F3'5'H. It may be that STR lost the function of F3'5'H and/or gained the function of F3'H in the process of variety improvement. We speculate that the weaker expression of *F3'5'H* in STR is due to the mutation of transcription factor or the gene itself. So, we compared the genomic region of *F3'5'H* in 'Strauss' and *C. persicum,* and there is no difference between them. In addition, we also analyzed the genomic region of *F3'Hs* in Strauss and *C. persicum,* and found that they have the same structures. Whether the mutation of transcription factors leads to changes in gene expression levels, which affects the production and accumulation of major anthocyanins, needs further research.

In this study, the recombinant plasmids pET21a-*STRF3'H1,* pET21a-*STRF3'H2a*, and pET21a-*STRF3'H2b* were constructed by using the prokaryotic expression vector of pET21a, and the optimal conditions for inducing proteins were determined. The expression levels of pET21a-*STRF3'H2a* and pET21a-*STRF3'H2b* were higher when induced at 0.5 mM IPTG at 20 °C for 16 hours, and the molecular weights of induced proteins were similar to those of the proteins encoded by the two genes. The recombinant protein pET21a-*STRF3'H1* has no protein band detected in the supernatant, but exists in the form of inclusion bodies. In the future, we could try to use the vector with fusion tag to improve soluble expression, or use the yeast eukaryotic system for protein expression.

Chapter 3 Identification of 5-*O*-glucosyltransferases involved in anthocyanin biosynthesis from *Cyclamen purpurascens*

3.1 Introduction

In different plants, the biosynthesis of anthocyanins undergoes different modifications, including glycosylation, methylation, acylation and so on (Yamazaki et al. 1999). Through the synergistic effects of these modifications, more diverse kinds of anthocyanins can be formed, so as to present a variety of floral colors. However, glycosylation always precedes other modifications, and it plays a constructive role in changing the hydrophilicity of anthocyanins, increasing their solubility and chemical stability, as well as facilitating their storage and transport in cells (Vogt and Jones, 2000). In plants, these glycosylation reactions are controlled by a specific family of glycosyltransferases (GTs). GTs is a multi-member family of transferases widely existing in organisms that catalyze the formation of glycosidic bonds between specific glycosylates and receptors (Meech et al. 2019). According to the similarity of sequence, conserved motifs and catalytic mechanism, GTs can be classified into 111 subfamilies (http://www.cazy.org/fam/acc_GT.html).

Uridine diphosphate (UDP)-dependent glycosyltransferases (UGTs), belong to the GT1 family, are a major class of enzymes that play a vital part in plant secondary metabolism. Flavonoid glycosylation is mainly catalyzed by UGTs. Plant UGT mainly uses small molecular compounds such as flavonoids, phenolic acids, terpenoids and plant hormones as glycosyl acceptors, and UDP-glucose, UDP-galactose or UDP-rhamnose as glycosyl donors (Lim et al. 2004; Bowleset al. 2006). The first UGT in plants was accidentally discovered by Nobel Laureate Barbara McClintock while studying the genetic instability of maize transposons. The *Bronze1* gene (X13500), which confers dark

pigmentation in corn kernels, was later shown to encode a flavonoid glycosyltransferase (UFGT) (Dooner & Nelson, 1977). UGTs have different regional selectivity in flavonoid glycation reactions. According to the different catalytic sites, UGT can be divided into 3-*O*-glycosyltransferase, 5-*O*-glycosyltransferase, 7-*O*-glycosyltransferase and the glycosyltransferase that catalyzes the generation of diglycosides. For example, anthocyanin 3',5'-*O*-glucosyltransferase (*UA3'5'GT*) cloned from the petals of *Clitoria ternatea* L., could catalyze the glycosylation of 3'-OH position and 5'-OH position of delphinidin 3-*O*-(6''-*O*-malonyl)-*β*-glucoside successively to produce delphinidin 3-*O*-(6''-*O*-malonyl)-*β*-glucoside-3',5'-di-*O*-*β*-glucoside (Kogawa et al. 2007). In addition, flavonoid *C*-glycosyltransferase has also been isolated from some plants (Noguchi et al. 2008; Brazier-Hicks et al. 2009; Falcone Ferreyra et al. 2013).

At present, the most studied glycosyltransferase gene is *3GT* gene, which is one of the key enzymes in the middle and downstream of anthocyanin biosynthesis pathway. Its function is to convert unstable anthocyanins into stable anthocyanins, and from colorless to colored, so that the maximum absorption spectrum shifts to the ultraviolet end. However, there are relatively few studies on *A5GT*, which catalyzes the glycosylation of anthocyanin at 5-*O*-position (Figure 3-1), and the anthocyanin-glucoside generated is more stable than mono-glycoside. Yamazaki et al. (1999) cloned cDNA of *A5GT* from *Perilla frutescens* for the first time by using mRNA differential display method. Since then, *A5GT* has been isolated from other plants: A5GT from *Dahlia* was reported in 2001, and revealed to have different substrate specificity with *Perilla* (Ogata et al. 2001); seven *A5GT* candidates were identified from *Gentiana trifloral*, of which only Gt5GT7 was verified to have A5GT enzyme activity (Nakatsuka et al. 2008); *A5GTs* were also isolated and characterized from monocotyledonous flower plants, *Iris hollandica* (Imayama T. et al. 2004), *Freesia hybrida*

(Ju et al. 2018). It can be concluded from the above relevant studies that the expression level of *A5GT* is correlated to the accumulation of major pigments in these plants, and A5GT from different plants has a wide range of substrate specificities. But as far as we know, A5GT of cyclamen has hardly been characterized.

Cyclamen is loved by people because of its rich ornamental trait, and is widely used in potted plants and garden greening. Most of commercial cultivars are obtained through the natural mutation and hybridization of wild *Cyclamen persicum*, they always have various color, shape, size, but no-good fragrance (Grey-Wilson 2002). Another wild species, *C. purpurascens*, is considered a precious material for the cultivation of aromatic cyclamen. However, in the early days, it was difficult to cross between *C. purpurascens* and *C. persicum* cultivars, because *C. purpurascens* always have very small and few flowers. Furthermore, histological observation showed that the fertilized ovule in *C. purpurascens* and *C. persicum* cultivars cross combination contained weak hybrid embryos without endosperm, and the endosperm eventually aborted (Ishizaka, 2018). To solved the problem of strong cross-incompatibility with *C. persicum* and other species, Ishizaka and Uematsu (1992) established the ovule cultivation system, which is a valuable method for creating interspecific *Cyclamen hybrids*. Subsequently, chromosome doubling technology overcomes seed sterility caused by lack of affinity between different genomes (Ishizaka and Uematsu 1995b). So far, *C. purpurascens* can really be used in the cultivation of fragrant cyclamen. Interestingly, all the F_1 progenies of the cross between *C. purpurascens* and *C. persicum* cultivars contain 3,5-diglucoside type anthocyanins in petals, which is same to *C. purpurascens* (the major pigment is Mv3,5dG) (Takamura et al. 2005). For example, F1 progenies of the cross between *C. purpurascens* and 'Strauss' have a pink slip with peonidin 3,5-diglucoside, cyanidin 3,5-diglucoside, and malvidin 3,5-diglucoside, and

a deep purple eye with malvidin 3,5-diglucoside, although the mainly pigment of 'Strauss' is Pn3Nh (Takamura et al. 2004). Fragrance cyclamen cultivar 'Kaori-no-mai' (KM), produced by crossing of *C. persicum* cultivars and *C. purpurascens*, accumulated Mv3,5dG as a dominant anthocyanin component and present purple-colored flower. While its ion beam-irradiated mutant 'Mayabi-no-mai' (MY), with red-purple-colored flower, accumulated malvidin 3-glucoside (Mv3G) (Ishizaka et al. 2012). It seems that the deletion of *A5GT* gene caused the change of flower color of MY and A5GT may make positive contribution to cyclamen flower color formation. Previously, three *5GT*-like genes (*Cpur5GT*s) have been isolated partially from *C. purpurascens*. Among them, *Cpur5GT2* was most likely to be a functional A5GT (Hase et al. 2012). This time we isolated the full-length ORF of *Cpur5GT2* and analyzed the enzyme activity *in vitro*.

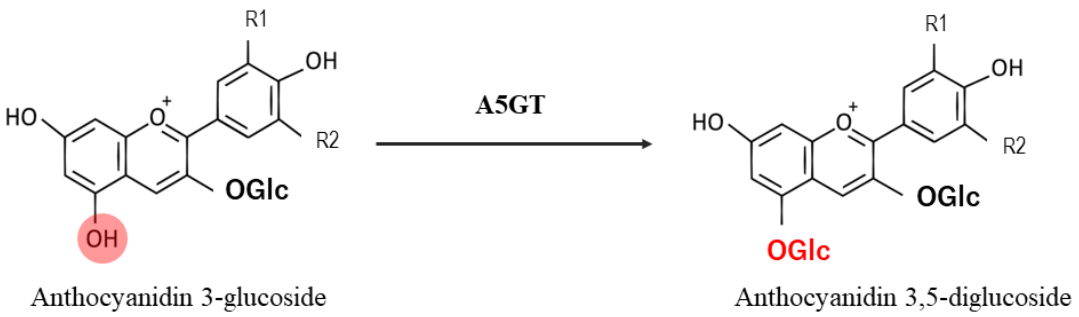


Fig. 3-1 Formation of 3,5-diglucoside type anthocyanin by anthocyanin 5-*O*-glucosyltransferase. R1, R2 could be substituted with -OH, -OCH3.

3.2 Plant materials

Cyclamen was grown in greenhouse at Saitama Institute of Technology, Japan. The petals, leaves, immature anthers, and petioles of *C. purpurascens* were collected and immediately frozen in liquid nitrogen then stored at - 80 °C until required.

3.3 Method

3.3.1 Extraction of genomic DNA

Genomic DNA was extracted from leaves of *C. purpurascens*. Detailed descriptions as 2.2.1.

3.3.2 Extraction of total RNA and synthesis of first-strand cDNA

Total RNA was extracted from different tissues (young petals, full-opened petals, immature anthers, leaves and petioles) of *C. purpurascens*. RNA extraction method and cDNA synthesis are detailed in 2.2.2 and 2.2.3.

3.3.3 Isolation of Cpur5GT *genes*

Based on the obtained gene fragments (*Cpur5GT1, Cpur5GT2* and *Cpur5GT3*) (Hase et al. 2012), 3' RACE and 5' RACE methods were carried out using the 5'/3'-RACE 2nd Generation Kit (Merck, Germany), to amplify unknown sequence upstream and downstream of *Cpur5GTs*. Gen-special primers were designed to obtained the full-length ORFs (Table 2). The amplified target gene fragments were subjected to gel recovery, ligated to pTAC-2 Easy vector (BioDynamics Laboratory Inc., Japan), transformed into *E. coli* JM109 competent cells, screened with blue and white spots, finally positive clones were picked for sequencing (Model 3500, Applied Biosystems) using the BigDye® Terminator ver. 3.1 Cycle Sequencing Kit (Applied Biosystems, MA, USA).

3.3.4 Bioinformatics analysis of Cpur5GT *genes*

Use the Basic Local Alignment Search Tool in NCBI to perform homologous sequence alignment of the amino acid sequences encoded by *Cpur5GT* genes. Predict the physical

and chemical properties and other information by online software ProtParam (http://web.expasy.org/protparam/). Protein the secondary structure prediction was carried out by Network Protein Sequence Analysis (Combet et al. 2000) using self-optimized prediction method with alignment (SOPMA) (Geourjon and Deléage 1995). The transmembrane regions of proteins encoded by *Cpur5GT1* and *Cpur5GT2* were predicted by online tool TMHMM (http://www.cbs.dt u.dk/services/TMHMM/). Multi-alignment analysis was performed by the Clustal W program (Thompson et al. 1994), the deduced amino acid sequence of Cpur5GTs were aligned with other A5GT proteins that obtained from the DDBJ/GenBank DNA databases. Phylogenetic trees were constructed using the Neighbor-Joining method (Saitou et al. 1987) with MEGA7 (Kumar et al. 2016).

3.3.5 Cpur5GT *genes expression analysis*

To investigate the expression levels of the isolated *Cpur5GT* genes in different tissues (young petals, full-opened petals, immature anthers, leaves and petioles), real-time PCR were performed on the Quant Studio™ 1 System with Power Up™ SYBR® Green Master Mix (Thermo Fisher Scientific). The reaction system in a total volume of 20 μl containing 100 ng of template cDNA, 10 μM of primer, 10 μl of 2 × Master Mix, and implemented under the standard cycling mode. The *eEF1α* genes were amplified as an internal control under the same conditions. Each experiment was done in 4 replicates, $2^{-\Delta\Delta CT}$ method was used to analyze the relative expression level of target genes (Livak and Schmittgen 2001). See 2.2.6 for specific operation steps.

3.3.6 Protein expression and purification

The plasmid pTAC2-*Cpur5GT2* with correct sequence and pET16b expression vectors are extracted and digested by the NdeI and BamHI. The digested products were recovered

and ligated at NdeI and BamHI restriction sites with ligation high (Toyobo, Japan). Transformed the recombinant expression vector pET16b-*Cpur5GT2* into *E. coli* strain BL21 (DE3) competent cells (Novagen). Picked the single colony of *E. coli* BL21 (DE3) embracing the reconstructed plasmids and cultured at 37 °C in LB liquid medium containing 100 μg ml^{-1} AMP untilled the OD_{600} was about 0.4-0.6, next induced the expression of the fused protein with 5 μM, 10 μM, 20 μM IPTG at 28 °C for 16 h. Recovered the expressed product, resuspended in Fast Break™ Cell Lysis Reagent (Promega) then purified the fusion protein utilizing the HisLink™ Spin Protein Purification System (Promega) in terms of the manufacturer's protocol. Used SDS-PAGE (15%) to detect the expression of the target protein, took the uninduced recombinant expression vector as a control. After determining the induction concentration of IPTG, perform a mass expression, see 2.2.10 for specific operations. Protein of mass expression was purified with Capturem ™ His-Tagged purification maxiprep columns.

Quantification of purified protein, refer to Bradford method (Bradford 1976). Prepare dye reagent (Bio-Rad protein assay, USA) in a volume of 1:4 dye reagent and deionized water, and filter through Whatman #1 filter. Prepare five BSA (Bio-Rad protein assay, USA) standards of different concentrations (0.078 mg ml^{-1}, 0.156 mg ml^{-1}, 0.39 mg ml^{-1}, 0.78 mg ml^{-1}) to make a standard curve. Pipet 100 μl of each standard and sample solution into a clean test tube. Add 5.0 ml of diluted dye reagent to each tube and mix. Incubate at room temperature for 10 min. Measure absorbance at 525 nm. Make a standard curve and calculate the sample concentration.

3.3.7 Enzyme assay of Cpur5GT2

The enzyme activity *in vitro* assay for *Cpur5GT2* was performed, the reaction mixture

consisting 35 μg purified protein, 2mM UDP-glucose or UDP-galactose as donor, 0.2 mM 3-glucoside type anthocyanidins (delphinidin 3-glucoside [Dp3G], cyanidin 3-glucoside [Cy3G], malvidin 3-glucoside [Mv3G], peonidin 3-glucoside [Pn3G], pelargonidin 3-glucoside [Pg3G]), petunidin 3-glucoside [Pt3G]) as acceptor, 100 mM Tris (pH 7.5), and the final volume was made up to 200 μl with pure water. For determine the Km value for UDP-glucose, kinetic parameters, all the reaction mixtures with a final reaction volume of 200 μl, were consisted 35 μg recombinant Cpur5GT2 protein and 100 mM Tris-HCl buffer (PH 7.5). For the measurement of the Km for acceptor substrate (Mv3G), each assay contained 2 mM UDP-glucose with different concentrations of Mv3G ranging from 0.1 to 1.0 mM. For determine the Km value for UDP-glucose, each assay contained 0.2 mM Mv3G and a UDP-glucose concentration that varied from 0.125 to 2 mM. After incubated the mixture at 28 °C for 1 h, 10% phosphoric acid was used to terminate the reaction. Then centrifugation of every mixture at 15,000 rpm for 5 min, the supernatant was filtrated through a 0.22 mm filter (Shimadzu, Japan) and subjected to high-performance liquid chromatography (HPLC) analysis. A Prodigy ODS-3 reversed-phase column (4.6×100-mm 3-μm 100Å, Phenomenex) was used to separate the metabolites at 30 ℃. The mobile phase consisted of 1.5% (v/v) phosphoric acid(A), 1.5% (v/v) phosphoric acid, 20% (v/v) acetic acid and 25% (v/v) acetonitrile solution (B). The elution program was proceeded over 60 min, at a flow rate of 0.3 ml min^{-1}. Quantify the reaction products by measuring the absorbance peak area at 525 nm.

Table 2. Primer used in Chapter 3

Primer	Sequence (5'-3')
Cpur5GT1-FP	ATGTATCCAAACAGTCCTTCGG
Cpur5GT1-FP1	CCAGAGGGCGGAGCCTAATCG
Cpur5GT1-FP2	TACTAGAATTTGCTTGGGGGC
Cpur5GT1-RP	TTATTTCTCTAGCAACACATTGACG
Cpur5GT1-RP1	TCCACAGAAATTGGTGGCTGC
Cpur5GT1-RP2	TACTTCCTCTTGTGCACACC
Cpur5GT1-RP4	ATTGAATATCTCATCGTCGGG
Cpur5GT1-RP5	CTTTCTCCATGAGAGCACGG
Cpur5GT2-FP	ATGGAGAATCGGTATCGTGTTC
Cpur5GT2-FP1	ACTTCAACGGCTATAGCGAGG
Cpur5GT2-FP2	AGCAACAAATGGAGGAGATCG
Cpur5GT2-RP	TTAATATTCGACAACACCTCTAATCTC
Cpur5GT2-RP1	GTGAGCACCATTCTACTATTTT
Cpur5GT2-RP2	CATTAGTGCTTTGGTCTGTCC
Cpur5GT2-RP3	TTATCTCCTCAACTGCTTCGG
Cpur5GT3-RP1	CTCCGAAAGAATTGTCCGACG
Cpur5GT3-RP2	AAGAAAGAGGGCAAGTCACG
NdeI *Cpur5GT2-FP*	CAT ATGGAGAATCGGTATCGTGTTC
BamHI *Cpur5GT2-RP*	GGATCC TTAATATTCGACAACACCTCTAATCTC
*eEF1a*Fw	CTGGTGGTTTTGAGGCTGG
*eEF1a*Rv	CTGGCCAGGGTGGTTCATGAT

3.4 Results and discussion

3.4.1 Cloning and biochemistry analysis of Cpur5GTs

On the basis of the existed gene fragments *Cpur5GT1*, *Cpur5GT2* and *Cpur5GT3*, we amplified their complete ORFs by RACE method. Sequencing results showed that *Cpur5GT1* and *Cpur5GT3* have identical nucleic acid sequences, and the original differences in the sequences of the two gene fragments was probably due to the mutation introduced by the Taq polymerase during PCR. So, we assumed two candidate genes *Cpur5GT1* and *Cpur5GT2*, they encoding 369 and 468 amino acid residues respectively. Using the CD-Search online tool in the NCBI website to analyze and predict the conserved domains of the protein encoded by Cpur5GT1 and Cpur5GT2, the results showed that Cpur5GT1 contains a typical glycosyltransferase domain, located between the 186 th and 303 th amino acids from the N-terminal, and belongs to the GTB-type glycosyltransferase superfamily; Cpur5GT2 also contains a typical glycosyltransferase domain, located between the 269 th and 395 th amino acids from the N-terminal, and belongs to the GTB-type glycosyltransferase superfamily (Figure 3-2a,b).

The physical and chemical properties of Cpur5GTs were analyzed by ProtParam (http://web.expasy.org/protparam/). The pI of Cpur5GT1 was 5.09, Cpur5GT2 was 5.0, the molecular weight of Cpur5GT1 was 41.45 kDa, while Cpur5GT2 had a similar molecular mass (52.96 kDa) as Dv5GT (53 kDa), Ph5GT (52 kDa) and Va5GT (54 kDa) (Ogata et al. 2001; Yamazaki et al. 2002; He et al. 2015).The instability index (II) of Cpur5GT1 and Cpur5GT2 were computed to be 37.29 and 33.42, respectively, both of them can be classified as a stable protein. The GRAVY of Cpur5GT1 and Cpur5GT1 was computed to be -0.0142, -0.283, respectively, that means they are hydrophilic protein. We also analyzed

the genomic structure of the *Cpur5GT1* and *Cpur5GT2*, the results show that there are no introns in their respective coding regions.

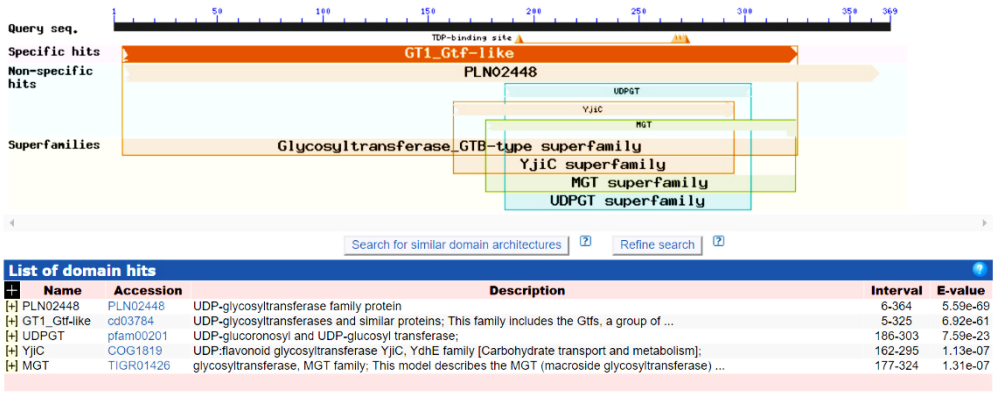
SOPMA predicted that the secondary structure of Cpur5GT1 is mainly made up of alpha helix (44.44%) and random coils (33.06%), while extended strands (14.63%) and beta turn (7.86%) only occupied a small part. Cpur5GT2 also have similar structural characteristics, composed of alpha helix (41.88%), random coils (35.68%), extended strands (15.81%) and beta turn (6.62%). The transmembrane structure of proteins encoded by *Cpur5GT1* and *Cpur5GT2* was predicted and analyzed by TM pred tool. And the results showed that the proteins encoded by the two genes did not have transmembrane regions (Figure 3-3a, b). When analyzed the transmembrane structure of Ph5GT (*Petunia hybrida*) and found that it is more similar to Cpur5GT2 (Figure 3-3c).

Comparison of the amino acid residues of Cpur5GT1 and Cpur5GT2, it only showed 24% identities (Figure 3-4). All aligned amino acid sequences had a conserved C-terminal domain of the UGT superfamily, which was known as plant secondary product glycosyltransferase (PSPG) box (Mackenzie et al.1997). The PSPG box is considered as the UDP-glucose binding region, composed of 44 amino acids, and the highly conserved amino acids are W_1, Q_4, L_8, H_{10}, $H_{19}CGWNS_{24}$, E_{27}, P_{39}, E_{43}/D_{43}, Q_{44} (numbered in PSPG box) (Caputi et al. 2012). Tryptophan (Trp, W) at position 22 can correctly locate UDP glucuronic acid, while arginine (Arg, R) can correctly locate UDP glucuronic acid. Serine (Sre, S) at position 23 is highly conserved in UDP glucuronic acid translocase (Shao et al. 2005). Notably, glutamine (Gln, Q) and histidine (His, H) are highly conserved as the last amino acid residue of the PSPG box in glucosyltransferase and galactosyltransferases, respectively (Kubo et al. 2004). Since the last amino acid in the PSPG domain of Cpur5GT1 and Cpur5GT2 was glutamine (Gln, Q), we conjectured that both of them preferred UDP-

glucose as the main glycation donor.

A neighbor-joining phylogenetic tree of the putative Cpur5GTs and several functionally characterized plant UGTs was present (Figure 3-5), it was obviously divided into three clusters upon the regiospecificity of UGTs for the aglycone substrate. The putative Cpur5GTs were clustered in two different branches, Cpur5GT1 and other 3GTs were gathered in Cluster I, Cpur5GT2 and other 5GTs, like Va5GT whose activities had been tested *in vitro*, were gathered in Cluster II (He. et al. 2015).

a.



b.

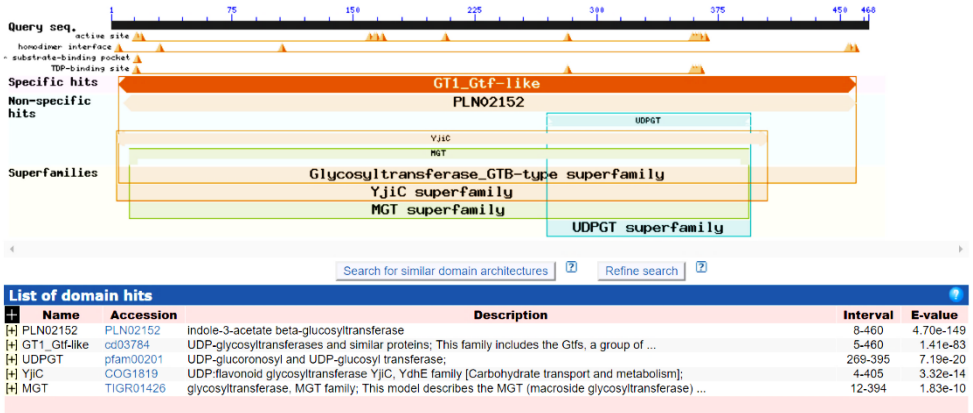


Fig.3-2 Predicts conservative structure domain of the protein encoded by *Cpur5GTs* (*a*)*Cpur5GT1* (*b*) *Cpur5GT2*

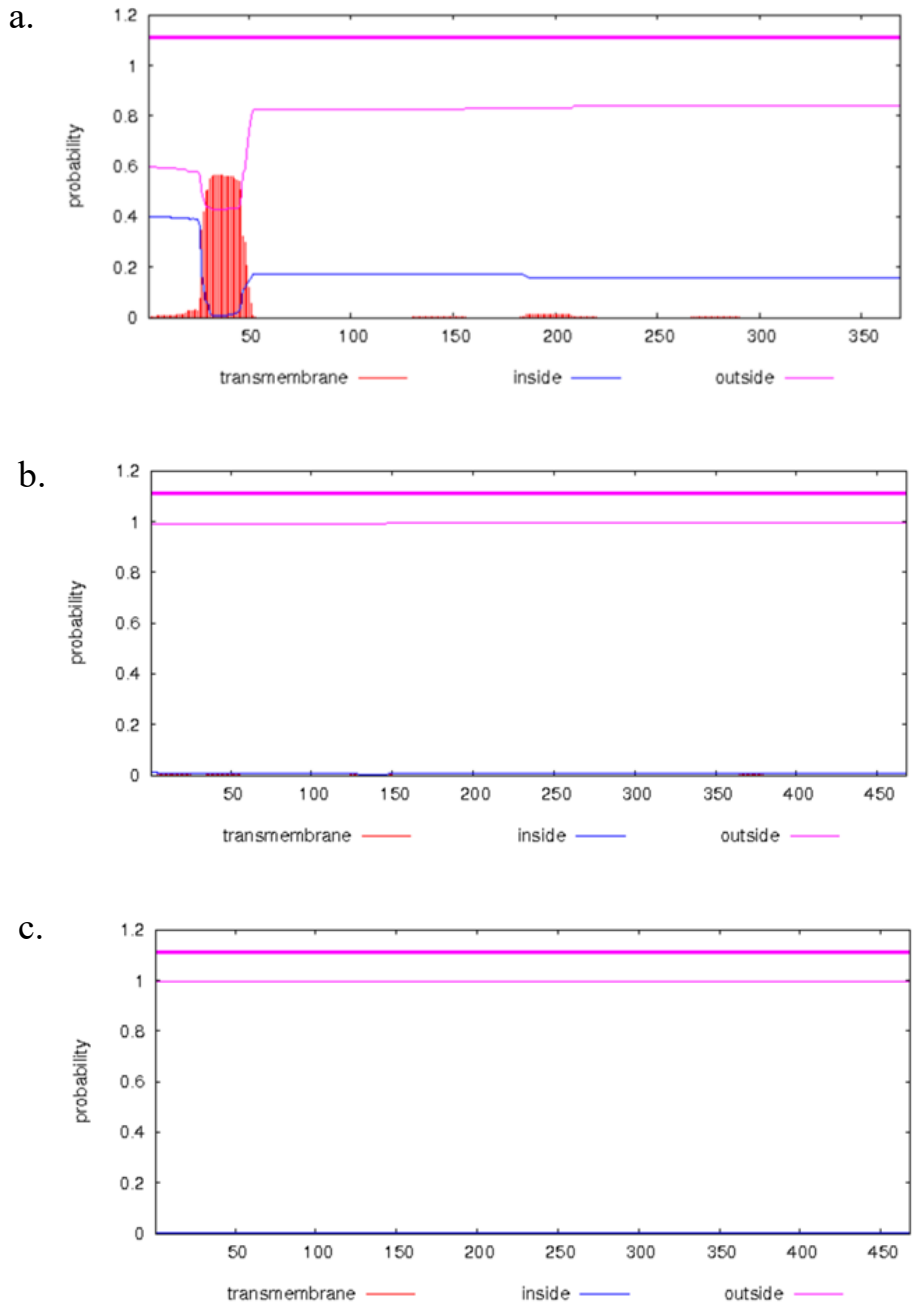


Fig. 3-3 Predicted transmembrane structure domain of (a) *Cpur5GT1*, (b) *Cpur5GT2*, (c) *Ph5GT* (*AB027455, Petunia hybrida*)

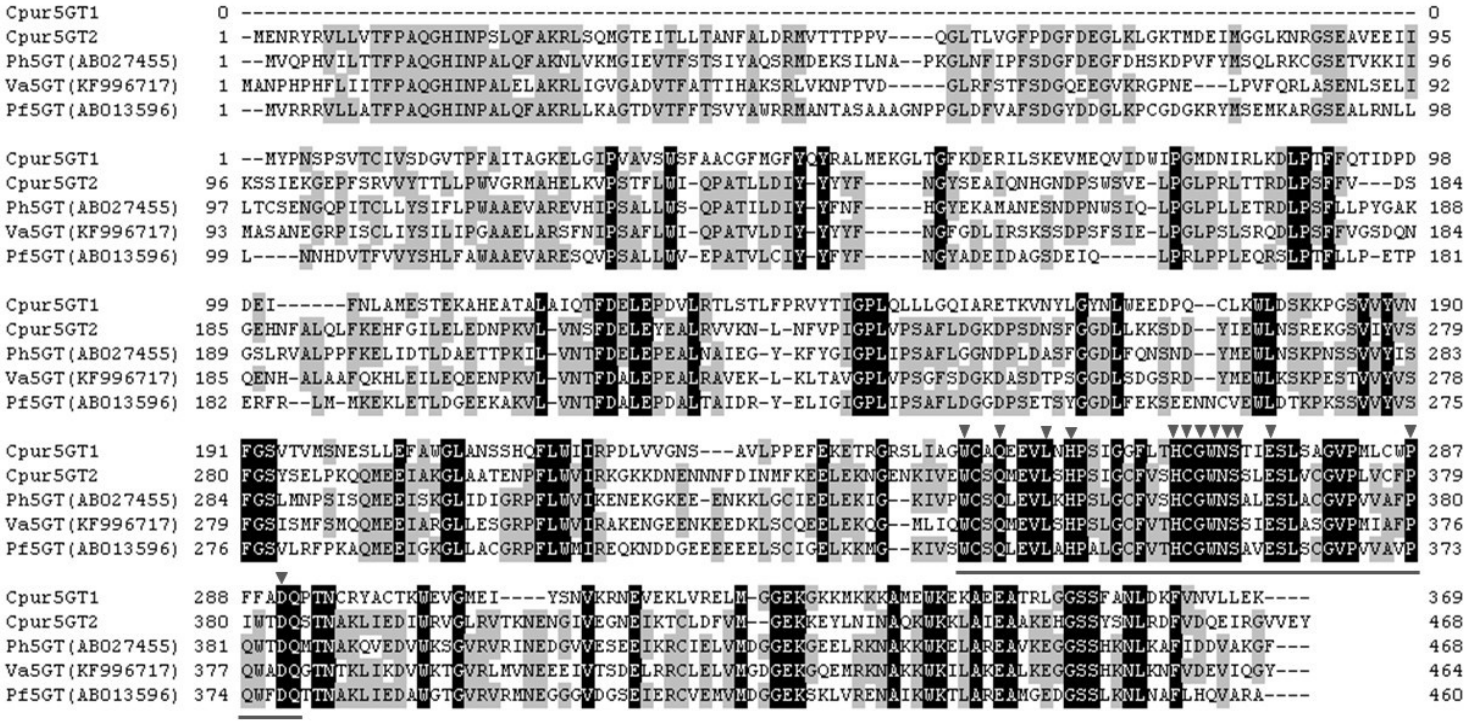


Fig.3-4 Alignment of the deduced amino acid sequences of Cpur5GT1, Cpur5GT2 and other published 5GTs. Identical amino acid residues are shaded in black, similar in grey. The PSPG box is underlined, the highly conserved amino acid residues in PSPG box are marked with triangles.

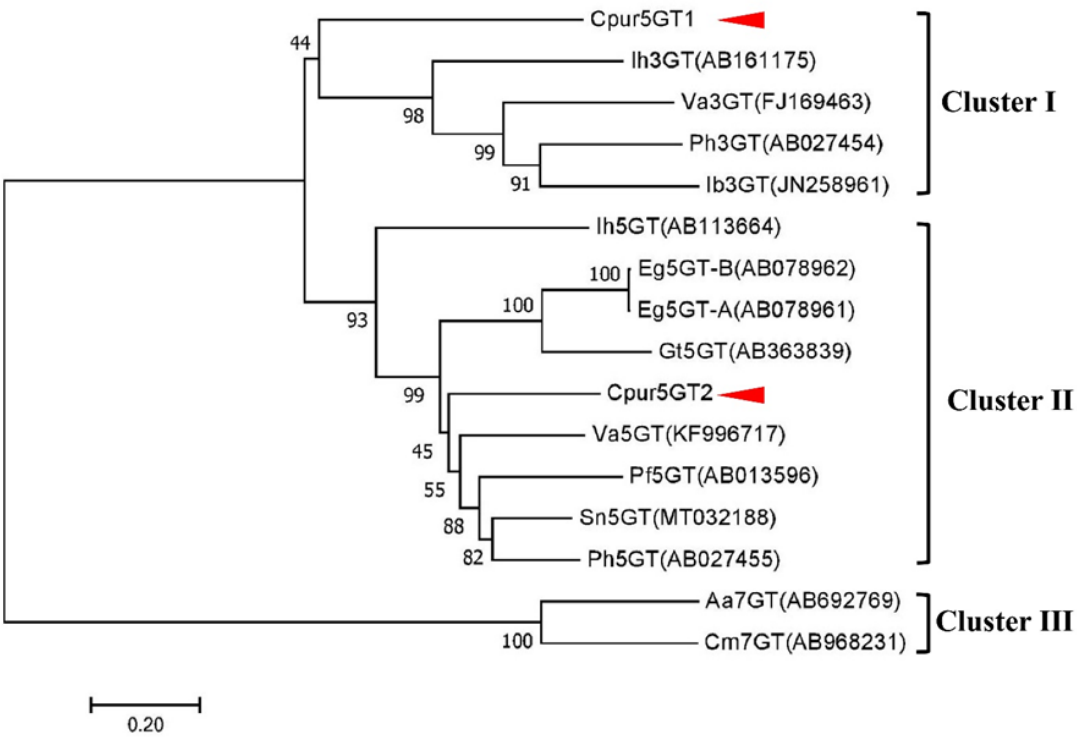


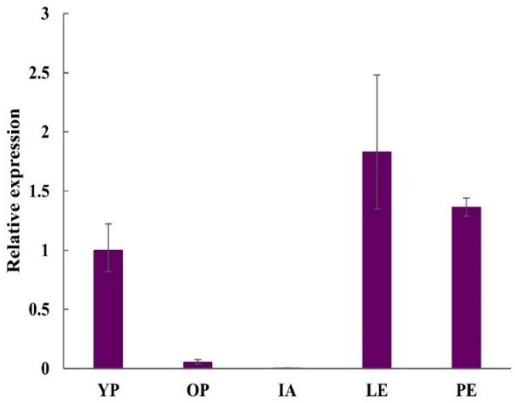
Fig. 3-5 Phylogenetic tree of Cpur5GT and GT members from other plant species. Three clusters were presented upon the regiospecificity of UGTs for the aglycone substrate. The putative Cpur5GT1 and other 3GTs were gathered in Cluster I, the putative Cpur5GT2 were located in Cluster II with other 5GTs. *Ih3G*(AB161175), *Iris hollandica*; *Va3GT*(FJ169463), *Vitis amurensis*; *Ph3GT*(AB027454), *Petunia x hybrida*; *Ib3GT*(JN258961), *Ipomoea batatas*; *Ih5GT*(AB113664), I*ris hollandica*; *Eg5GT-A*(AB078961), *Eustoma grandiflorum*; *Eg5GT-B*(AB078962), *Eustoma grandiflorum*; *Gt5GT7*(AB363839), *Gentiana trifloral*; *Va5GT*(KF996717), *Vitis amurensis*; *Pf5GT*(AB013596), *Perilla frutescens*; Sn5GT(MT032188), *Solanum nigrum*; Ph5GT(AB027455), *Petunia x hybrida*; Aa7GT(AB692769), *Agapanthus africanus*; *Cm7GT(*AB968231), *Campanula medium*

3.4.2 Expression pattern of Cpur5GT *genes*

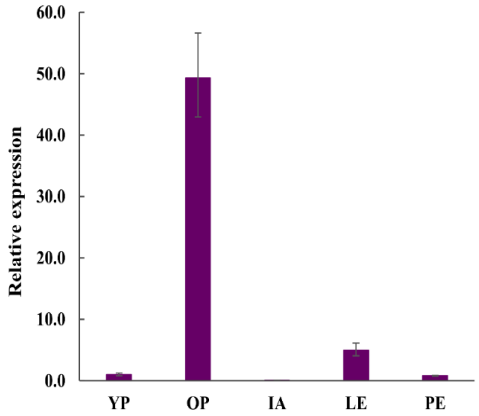
The expression profile of two *Cpur5GT*s in young petals (the paler pigmented stage), full-opened petals, immature anthers, leaves and petioles were surveyed by real-time PCR, two genes expressed scarcely in immature anthers, significantly different in other tissues. Higher transcripts of *Cpur5GT1* were observed in leaves, petioles, young petals, and lower in full-opened petals (Figure 3-6a). Whereas, *Cpur5GT2* expressed most strongly in full-opened petals, 10 times that of the leaves and 50 times that of the young petals (Figure 3-6b). The expression patterns of some genes that related to flower coloration change with the development of flowering. For example, Akita et al. (2018) have identified a functional co-pigmentation-related *FLS* gene from *C. purpurascens* (*CpurFLS2*). This gene strongly expressed at the early stage of flower development, and reduced expression in full opened flowers. While, *Cpur5GT2* was observed strong expression at later stage and weak expression at early stage of flower development (Figure 3-6b).Prior studies pointed out that the expression of *A5GT* was closely related to the accumulation of bis-glycosidic anthocyanins: the transcription levels of *Va5GT* in berry skins of *Vitis amurensis* intensity increased with the cumulation of their mainly anthocyanins (He et al. 2015); in *Freesia hybrida*, *Fh5GT1* and *Fh5GT2* were expressed in all the stages of flower development, and got the highest expression level when petals were completely pigmented (Ju et al. 2018). According to the classification of genes involved in anthocyanin biosynthesis in *Gentiana 3GT* and *5AT* that catalyze the later steps of anthocyanin biosynthesis belong to group III (Nakatsuka et al. 2005). Group III were not expressed or expressed at a very low level at the first stage of flower development, but their expression increased with flower development before the flower got full opened. *Cpur5GT2* was responsible for downstream modification in the anthocyanin synthesis pathway, its expression pattern in floral organs

was similar to that of *3GT* and *5AT* in *Gentiana*. Moreover, the tissue-specific expressions of *Cpur5GT2* resembled *Fh5GTs* expressed higher in pigmented petals than leaves and other tissues (Ju et al. 2018).

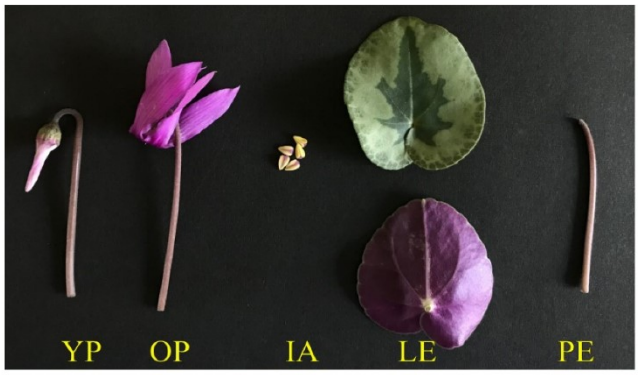
Combined with gene expression patterns and the bioinformatic analysis of the deduced amino acid sequences, we more suspected that *Cpur5GT2* rather than *Cpur5GT1* took a part in flower coloration in *C. purpurascens*.



(a) Expression pattern of *Cpur5GT1*



(b) Expression pattern of *Cpur5GT2*



(c) Different tissues of *C. purpurascens*

YP: young petals OP: full-opened petals IA: immature anthers

LE: leaves PE: petioles

Fig. 3-6 Expression analysis of *Cpur5GTs* in different tissues

3.4.3 Expression of recombinant Cpur5GT2 in vitro

The recombinant plasmid pET16b-*Cpur5GT2* was introduced into *E. coli* BL21 (DE3) cells and expressed under IPTG induction to determine the *in vitro* activity. SDS-PAGE analysis showed that the molecular mass of the crude protein induced at 5 μM,10 μM, 20 μM IPTG was all approximately equal to the expected molecular weight of the Cpur5GT2 protein (52 kDa). Finally, 5 μM IPTG was determined as the optimal concentration for inducing recombinant soluble Cpur5GT2 protein, and got a purified target protein band (Figure 3-7). Expanded cultivation *E. coli* BL21 (DE3) embracing the reconstructed plasmids pET16b-*Cpur5GT2* and induced with 5 μM IPTG at 28 ℃ for 16 h. In order to calculate the concentration of the target protein, we took different concentrations of BAS as samples, measured the absorbance at 525 nm, and made a standard curve $y=0.9283x+0.0537$ ($R^2=0.9993$). From this calculation, the purified target protein concentration is 0.523 mg ml^{-1}.

3.4.4 Enzyme assay of recombinant Cpur5GT2

To determine the acceptor specificity, UDP-glucose was used as the glycosyl donor, six potential substrates were used as acceptors to react with the purified recombinant Cpur5GT2 protein, and the products were detected by HPLC. When incubated with six kinds of anthocyanidin 3-glucoside as substrates, additional peaks were formed, which were identified as their corresponding anthocyanidin 3,5-diglucoside based on the retention time and UV spectrum of standard samples (Figure 3-8). Especially for Mv3G, Cpur5GT2 showed a higher relative activity (Table 3). Previous reports indicated that A5GTs always exhibited the maximum enzymatic activity to the precursors of their primary anthocyanins. For *Perilla frutescens*5GT, Cy3G was the most preferable substrate than other examined,

as cyanidin-derived anthocyanins were mainly cumulated in its red leaves (Yamazaki et al. 1999). Since cyanidin-, pelargonidin- derived anthocyanins were the main pigment in *Dahlia variabilis*, Dahlia5GT showed a higher affinity for Cy3G and pelargonidin 3-glucoside (Pg3G), but very low relative activity for Dp3G that has not been detected in Dahlia yet (Ogata et al. 2001). In our study, despite cyanidin 3,5-diglucoside (Cy3,5dG), delphinidin 3,5-diglucoside (Dp3,5dG), peonidin 3,5-diglucoside (Pn3,5dG), pelargonidin 3,5-diglucoside (Pg3,5dG) and petunidin 3,5-diglucoside (Pt3,5dG) have not been identified in *C. purpurascens*, Cpur5GT2 can still catalyze glycosylation at the 5-hydroxyl position of their 3-*O*-glycoside precursors, and has higher relative activity to Dp3G and Cy3G (Table 2). These indicated that Cpur5GT2 has broad substrate specificity. The mainly anthocyanins in the slips of the hybrids of *C. purpurascens* × *C. persicum* cultivars is 3,5-diglucoside type, and some scholars have pointed out that the glycosylation of anthocyanins in the slips may be caused by genes in *C. purpurascens* genome (Takamura et al. 2005; Yamazaki 2018). We speculated that it is precisely because of the broad substrate specificity of *Cpur5GT2* that hybrids of *C. purpurascens* × *C. persicum* cultivars have different 3,5-diglucoside types of anthocyanins. However, the acceptor substrate specificity of Cpur5GT2 was different with Va5GT (*Vitis amurensis* Rupr. cv. 'Zuoshanyi' grape), which preferentially glycosylated *O*-methoxylated anthocyanins (Mv3G, Pt3G, Pn3G) rather than their hydroxylated counterparts (Dp3G, Cy3G) (He et al. 2015). Many scholars have also argued that UGTs involved in plant secondary metabolism always have higher regiospecificity or regioselectivity to specific positions of hydroxyl groups, rather than the structure of sugar acceptors (Sun et al. 2016; Vogt and Jones 2000).

For sugar donor specificity analyses, UDP-glucose and UDP-galactose were tested as glycosyl donors respectively, with Mv3G as acceptor. However, the results showed that UDP-galactose could not serve as a sugar donor (Figure 3-9), which was consistent with

our prediction and was similar to the capability of Dv5GT and Va5GT (Ogata et al. 2001; He et al. 2015). The *Km* value for Mv3G was 79.2 μM against UDP-glucose as the glycosyl donor, and the *Vmax* value for M3,5dG formation was 18.6 nM min^{-1} mg^{-1}. The apparent *Km* value for UDP-glucose of 454 μM and a *Vmax* value of 18.1 nM min^{-1} mg^{-1} were obtained. These results are comparable with *Km* values of Va5GT (80.9 μM for Mv3G and 0.213 mM for UDP-glucose, respectively) (He et al. 2015).

Anthocyanin biosynthesis is a complex process and each enzyme does not act independently, the relative concentration of potential substrates may be one of the most crucial factors determining its activity in plants. Just as Pt3G and Pg3G that could be served as substrate for Cpur5GT2 *in vitro*, but no petunidin-derived anthocyanins have been detected in *C. purpurascens*, *C. persicum* cultivars and their interspecific hybrids (Ishizaka 2018). Of course, it cannot be excluded that there are other A5GTs in *C. purpurascens*, which have different relative activities to various anthocyanin substrates, since UGTs in flavonoid biosynthesis are usually encoded by a multi-gene family in many plants (Yonekura-Sakakibara and Hanada 2011). The above results revealed that Cpur5GT2 exhibited the activity for transferring the glucose moiety from UDP-glucose to the 5-position of anthocyanidin 3-glucoside, and displayed a broad substrate specificity.

To determine the function of Cpur5GT2 in detail, we will analyze the substrate specificity of Cpur5GT2 using other flavonoids such as flavonols. Moreover, to reveal the relationship between Cpur5GT2 and flower-coloration in cyclamen, the analysis of *Cpur5GT2*-like gene in the *5gt* mutation (ex. MY, Ishizaka et al. 2012) will also be explored.

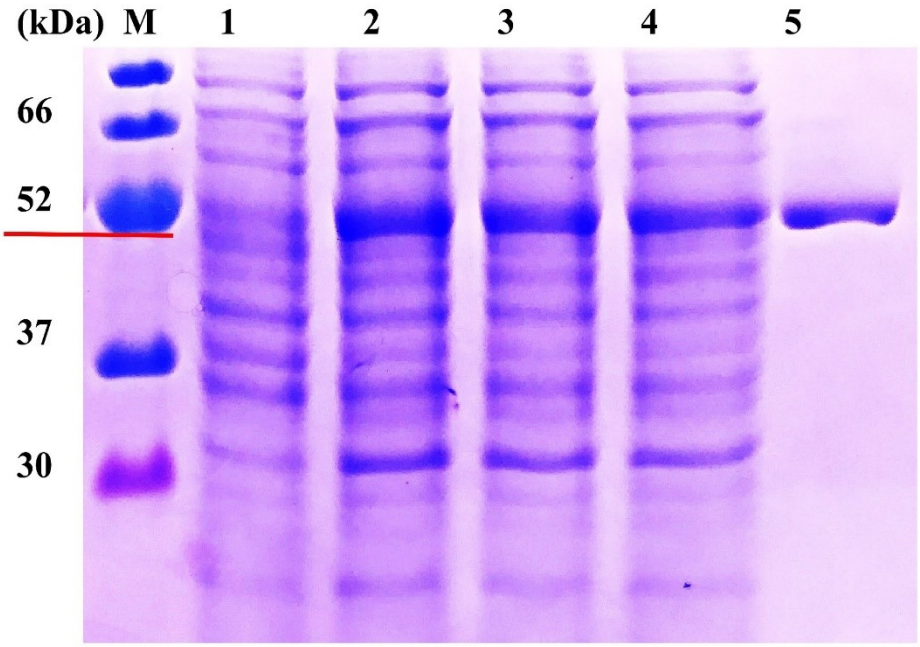
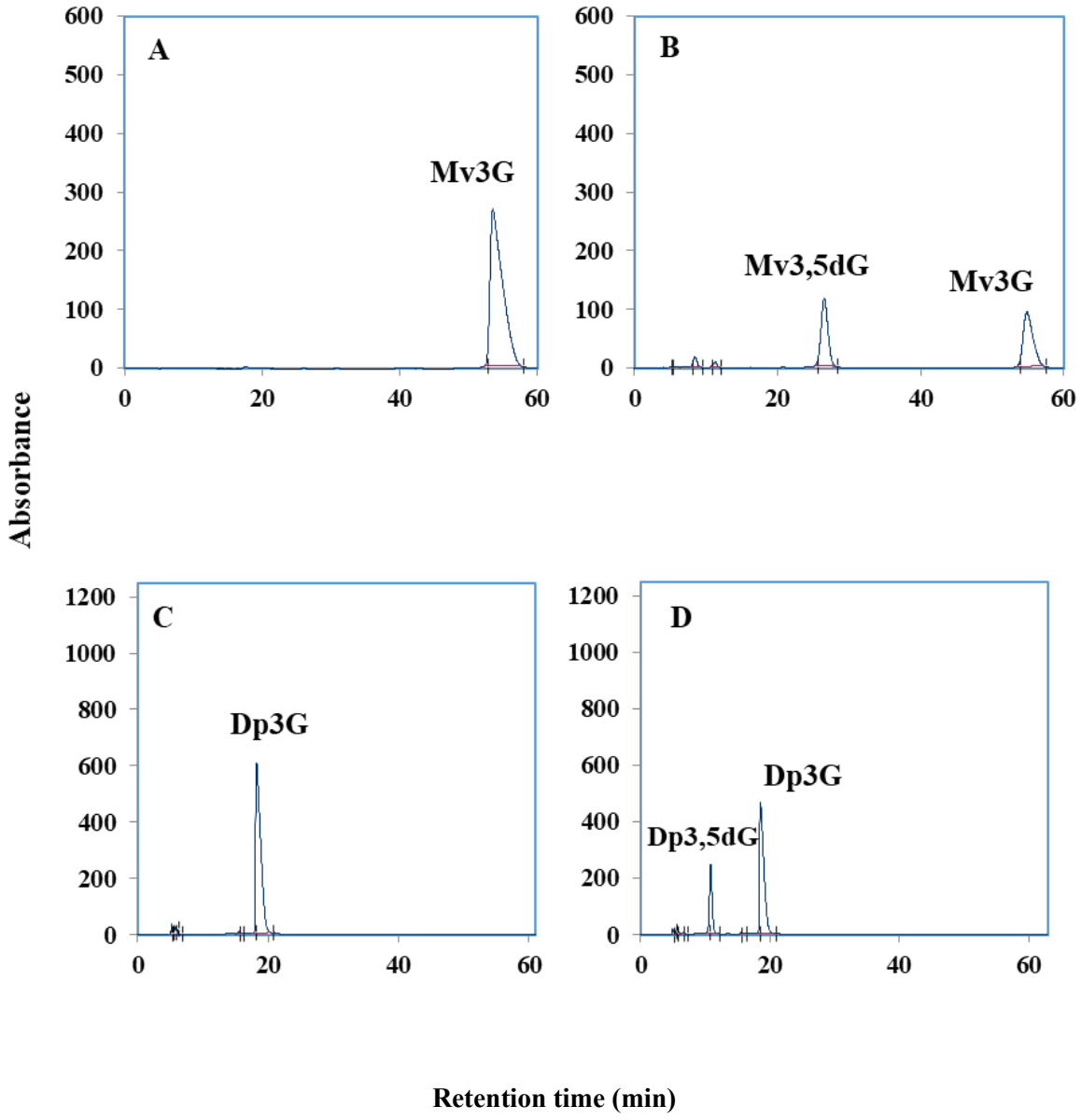
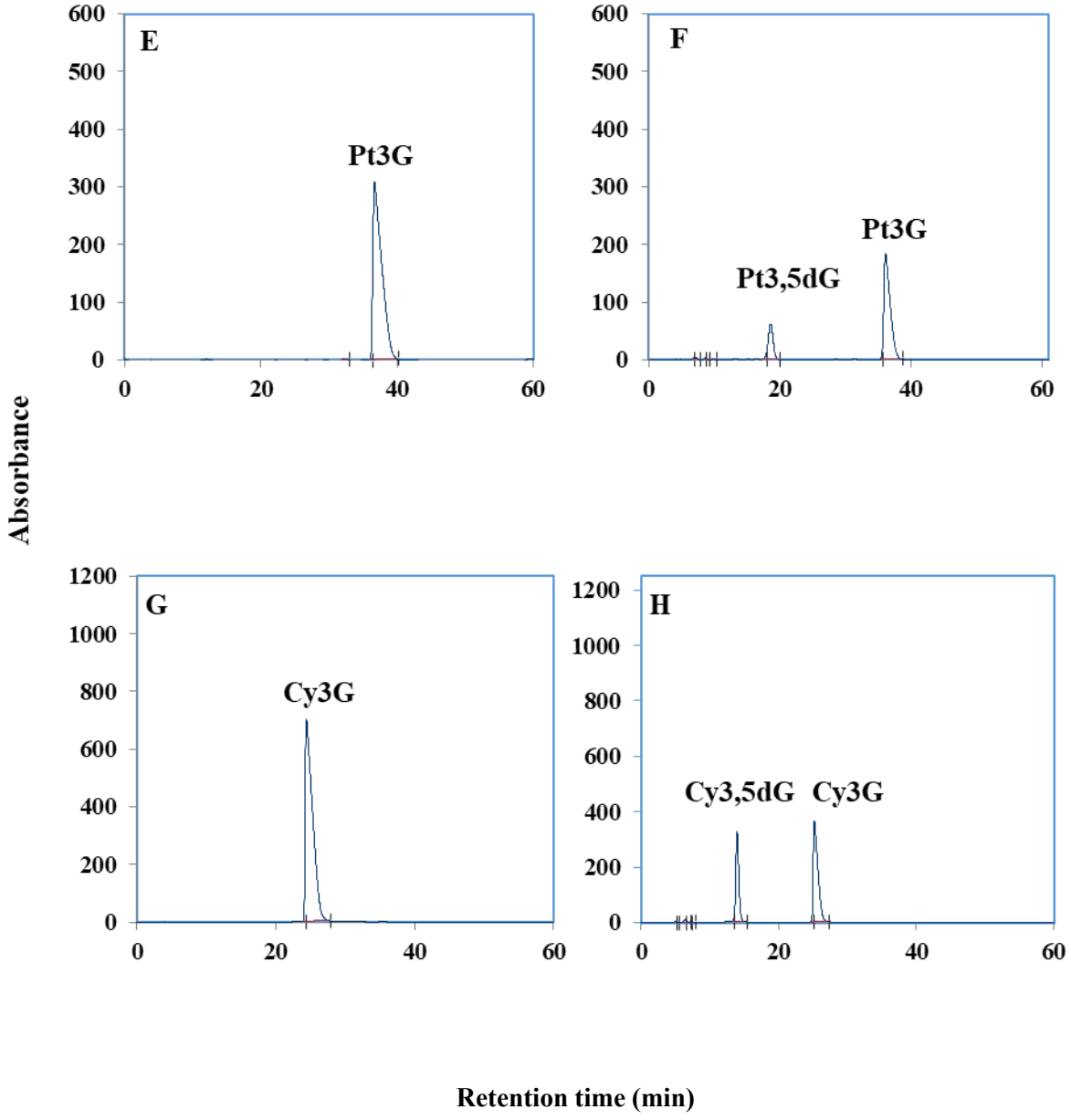


Fig. 3-7 SDS-PAGE analysis of the recombinant pET16b-*Cpur5GT2*

Lane M: molecular mass marker; lane1, 0 IPTG; lane 2, 5 μM IPTG; lane 3, 10 μM IPTG; lane 4, 20 μM IPTG; lane 5, purified recombinant protein; Induction condition: 28 ℃, 16 h.


A
Mv3G
B
Mv3,5dG
Mv3G
C
Dp3G
D
Dp3G
Dp3,5dG
Absorbance
Retention time (min)


E
Pt3G
F
Pt3,5dG
Pt3G
G
Cy3G
H
Cy3,5dG
Cy3G
Absorbance
Retention time (min)

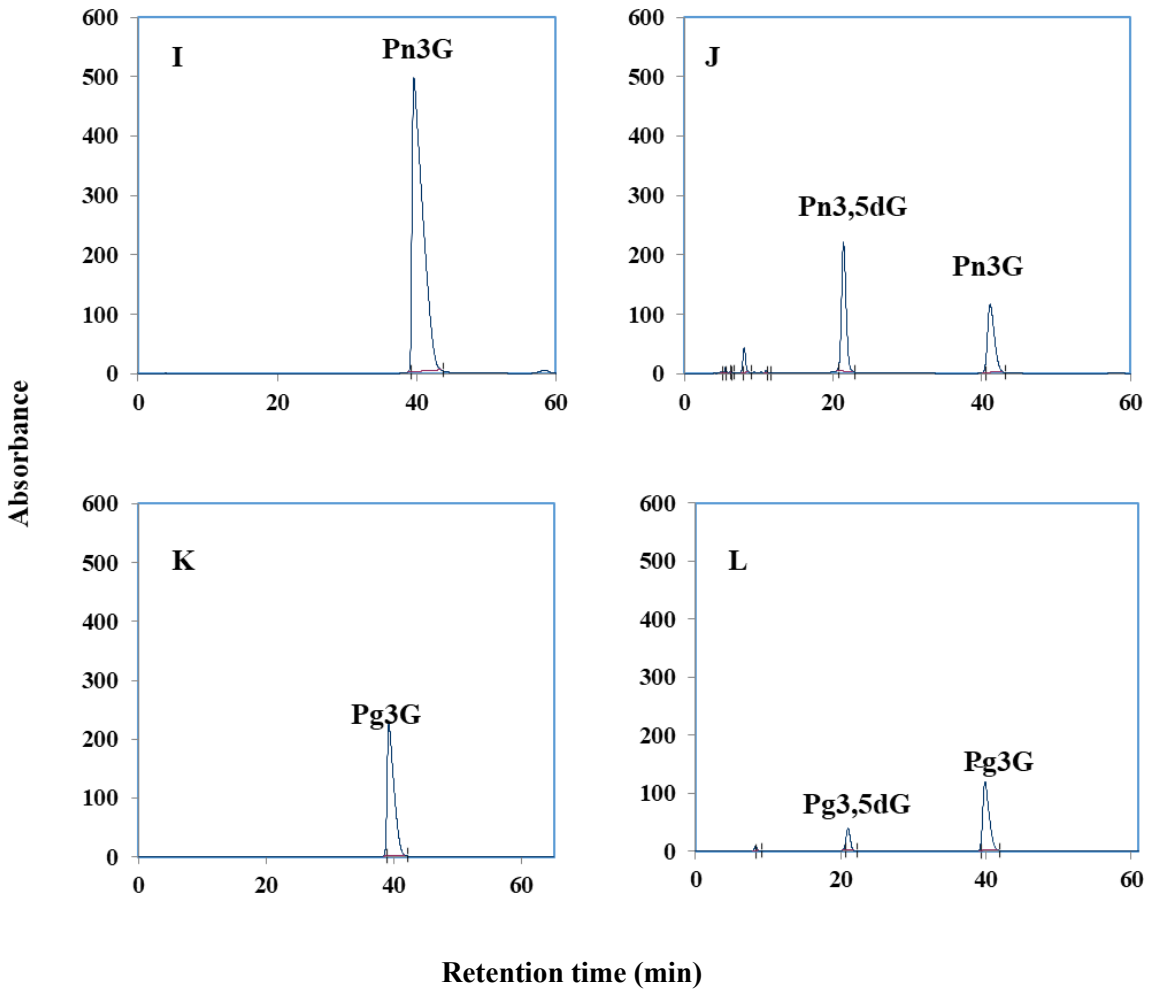


Fig. 3-8 HPLC analysis of Cpur5GT2 reaction products obtained from the *in vitro* enzyme assay. (A-L) Six potential substrates including Mv3G (A, B), Dp3G (C, D), Pt3G (E, F), Cy3G (G, H), Pn3G (I, J) and Pg3G (K, L) were incubated with (B, D, F, H, J, L) or without (A, C, E, G, I, K) the purified Cpur5GT2.

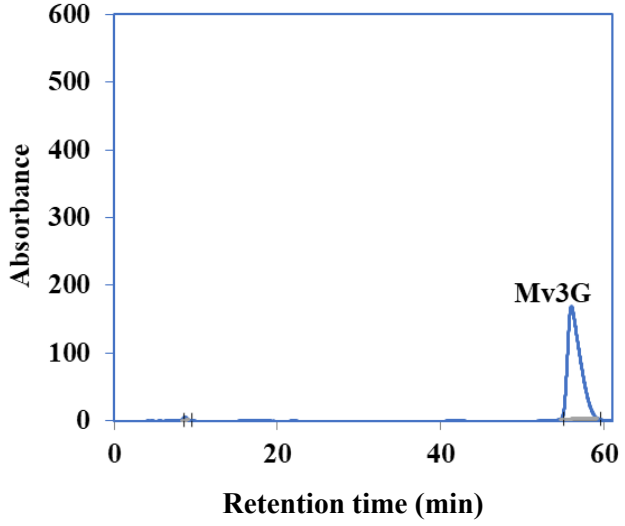


Fig. 3-9 HPLC analysis of Cpur5GT2 reaction product obtained from the *in vitro* enzyme assay. The reaction used Mv3G as the acceptor substrate and UDP-galactose as the sugar donor.

Table 3. Relative activity of Cpur5GT against several potential substrates

Substrate	Product	Relative activity (%)
Mv3G	Mv3,5dG	100[a] ± 2.2
Dp3G	Dp3,5dG	86.8 ± 3.3
Pt3G	Pt3,5dG	18.8 ± 2.3
Cy3G	Cy3,5dG	78.3 ± 12.4
Pg3G	Pg3,5dG	21.6± 3.9
Pn3G	Pn3,5dG	72.4 ± 7.1

[a] Relative activity analyses were performed with UDP-glucoside as the donor. The amount of reaction product was calculated from the area of the HPLC chromatogram recorded at 525 nm. The activity of Cpur5GT2 against malvidin 3-glucoside (Mv3G) was considered 100 %, which was used to determine its relative activity against other substrates.

3.5 Concluding remarks

This study has confirmed the complete ORF of *A5GT* from cyclamen for the first time. *Cpur5GT1* encoding 369 amino acid residues and the molecular weight of Cpur5GT1 was 41.45 kDa. *Cpur5GT2* encoding 468 amino acid residues, the molecular mass is 52.96 kDa, which is similar to the molecular weight of A5GT of other plants with established functions. Sequence alignment shows that they all have the UGT superfamily conserved PSPG domain. However, phylogenetic analysis shows that Cpur5GT1 and 3GTs /UFGT from other plants are clustered together, while Cpur5GT2 and other A5GTs are clustered together. Therefore, we preliminarily infer that Cpur5GT2 is more likely to be a typical A5GT. In order to clarify the expression pattern of *Cpur5GTs*, we compared the transcription of *Cpur5GT1* and *Cpur5GT2* in different tissues. The results showed that Cpur5GT1 was expressed in large amounts in leaves and young petals, but very low in full-opened petals. However, the expression of *Cpur5GT*2 was highest in full-opened petals, followed by leaves, and its expression trend was consistent with the accumulation level of anthocyanins in various tissues. Previous studies have also shown that the expression of *A5GT* is related to the accumulation of anthocyanins in plant tissues. In summary, we further infer that *Cpur5GT2* encodes a typical A5GT. Therefore, the recombinant prokaryotic expression vector pET16b-*Cpur5GT2* was constructed and the enzyme activity was detected *in vitro*. By adjusting the induction time, IPTG concentration and induction temperature, the optimal induction conditions for the recombinant protein were determined, 50 μM of IPTG, 28 °C,16 h. A large amount of induction and purification were carried out under the above conditions, and then the obtained protein was subjected to enzyme assay. Taking six kinds of 3-glucoside type anthocyanin as the substrates and UDP-glucose as the sugar donor, the catalytic reaction was carried out, and the reaction products were detected by HPLC. The

results demonstrate that Cpur5GT2 could catalyze glycosylation of all substrates at 5-*O*-position, and showed the highest relative activity to Mv3G, which may be because Mv3,5dG is the main pigment in *C. purpurascens.* However, UDP-galactose cannot be used as a glycosyl donor for the glycosylation catalyzed by Cpur5GT2. These results laid the foundation for further research on the molecular mechanism of glycosylation modification at the end of the cyclamen anthocyanin metabolic pathway.

Chapter 4 Conclusions

In this study, *F3'H* and *A5GT* involved in cyclamen anthocyanin synthesis have been characterized for the first time. This genetic information paved the way for comprehensively revealing the molecular mechanism of cyclamen anthocyanin synthesis and also provided a theoretical reference for breeding novel cyclamen varieties by genetic engineering in the future. The following conclusions can be drawn from the results of the experiment studies:

1. Three full-length ORFs (*STRF3'H1*, *STRF3'H2a*, *STRF3'H2b*) of *F3'Hs* of cyclamen were obtained by homologous cloning and RACE method, and the base sequence similarity of *STRF3'H2a* and *STRF3' H2b* reached 99.4%.

(1) By comparing their DNA sequences with cDNA sequences, it was found that *STRF3'H1* contained one intron, *STRF3'H2a* and *STRF3'H2b* each contained two introns. The deduced amino acid sequences of these candidate genes were compared with F3'H amino acid sequences from other plants, and they presented the same active sites and conserved domains. However, the prediction of transmembrane structure showed that STRF3'H2a and STRF3'H2b each had a transmembrane region, F3'H1 did not.

(2) The results of real-time PCR showed that the expression trends of the three candidate *STRF3'H* genes during flower opening were consistent, all of them had the highest expression level at the early stage of flower opening (non-staining stage), then gradually declined, and the lowest expression level at the full-opened stage. In addition, the expression of *STRF3'Hs* was detected in leaves, and the expression level was higher than that at the full-opened. These indicated that STRF3'Hs are related to the formation and accumulation of anthocyanin precursors, and plays a role in the secondary metabolic pathways other than anthocyanins.

(3) According to the real-time PCR analysis, *F3'Hs* were expressed stronger in STR and weaker in *C. persicum*. On the contrary, *F3'5'H* expressed stronger in *C. persicum* than in STR. We compared the genomic structure of the *F3'H* genes and of *F3'5'H* gene between STR and *C. persicum*, but found no difference. It suggested that the differences in expression of these two genes in STR and *C. persicum* may not be caused by mutations in the genes themselves. Therefore, whether the gene expression level is decreased due to mutations of the related transcription factors requires further investigate.

(4) In order to study the functions of the three candidate genes, we chose pET21a as prokaryotic expression vector, constructed recombinant plasmids, determined the conditions of protein induction, and successfully induced the expression of the corresponding protein in *E. coli* BL21(DE3). SDS-PAGE detection revealed that the induced pET21a-*STRF3'H1* recombinant protein mainly exists in the form of inclusion bodies. The induced pET21a-*STRF3'H2a* and pET21a-*STRF3'H2b* recombinant proteins exist in the form of soluble protein. These have laid a foundation for future research on the substrate specificity of F3'H and revealing the role of F3'H in anthocyanin biosynthesis.

2. A5GT is one of the key enzymes responsible for the modification of anthocyanins, and it is responsible for catalyzing the glycosylation reaction at the 5-*O*-position of anthocyanins. In this study, we used the wild fragment cyclamen, *C. purpurascens* as the material, and obtained two full-length ORF of *A5GT*, *Cpur5GT1* and *Cpur5GT2*.

(1) Bioinformatics analysis of the amino acid sequences encoded by the two genes revealed that both of them shared the PSPG conservative domain specific to UDPG. However, the constructed phylogenetic tree showed that Cpur5GT2 clustered with other known functions of A5GT, while Cpur5GT1 clustered with 3GT of other plants.

(2) *Cpur5GT1* and *Cpur5GT2* were almost not expressed in the immature anthers of *C. purpurascens*, but expressed in other tested tissues, and the expression levels were significantly different. The expression pattern of *Cpur5GT2* was more consistent with that functional *A5Gs* from other plants. They were all developmentally regulated and have tissue specificity and are related to the biosynthesis of anthocyanins. Besides, we also detected expression of *Cpur5GTs* in the leaves, possibly because they were involved in the glycosylation of other secondary metabolites in the leaves. Considering the analysis of the deduced amino acid sequence, phylogenetic relationship and expression patterns, *Cpur5GT2* is more likely to encode a typical A5GT.

(3) To analyze the function of Cpur5GT2 *in vitro*, we constructed the expression vector pET16b-*Cpur5GT2* and determined the optimal induction conditions for protein induction. When UDP-glucose was used as the sugar donor, the recombinant protein pET16b-*Cpur5GT2* could respectively catalyze the glycosylation of six 3-*O*-glucoside type anthocyanins to generate their corresponding 3,5-diglucoside anthocyanins. And the recombinant protein showed the strongest catalytic ability to Mv3G, which is speculated to be related to the fact that Mv3,5dG is the main pigment in *C. purpurascens*. But when UDP-galactose was used as the sugar donor, the glycosylation reaction cannot proceed.

Outlook

This time, three candidate *STRF3'H* genes were isolated, but whether they all encoding the functional F3'H needs to be verified. The expression of *F3'H*, *F3'5'H* in 'Strauss' and *C. persicum* differed significantly, this study conjectured that it may be due to the mutation of the related transcription factors, and this speculation still needs further analysis and confirmation. *STRF3'Hs* were weakly expressed in *C. persicum* but strongly expressed in 'Strauss'. We inferred that F3'H is associated with 'Strauss' flower color mutation, but the mechanism that affects flower color formation needs to be explored in depth. Furthermore, flower color formation is a complex system, involving many enzymes and genes, which restrict and interact with each other. Next, the differences of other genes involved in anthocyanin biosynthesis between 'Strauss' and *C. persicum* should be analyzed in detail to systematically clarify the reasons for red cyclamen flower color mutation.

The present study demonstrated that the recombinant Cpur5GT2 could catalyze 5-*O*-glycosylation of 3-glucoside type anthocyanins *in vitro*. In the future, *in vivo* characterization of Cpur5GT2 should be performed to reveal the role of Cpur5GT2 in anthocyanin biosynthesis in plants. For example, Virus Induced Gene Silencing (VIGS) technology can be used to specifically silence the *Cpur5GT2* gene in *C. purpurascens*, and then the flower phenotype can be detected and analyzed to clarify the function of this gene. In addition, it has been shown that in many plant species, the UGTs involved in flavonoid biosynthesis are usually encoded by a multigene family, so whether there are other functional A5GTs in *C. purpurascens* is also a topic for us in the future.

As genes involved in anthocyanin biosynthesis of cyclamen continued to be cloned and characterized, our understanding of the molecular mechanism of flower color formation will become more thorough. And these genes will also be served as the useful tools for the

targeted regulation of gene expression and the targeted modification of cyclamen flower color by molecular biotechnology.

References

Akita Y, Kitamura S, Mikami R, Ishizaka H (2018) Identification of functional *flavonol synthase* genes from fragrant wild cyclamen (*Cyclamen purpurascens*). *J Plant Biochem Biotechnol* 27:147-155

Akita Y, Ishizaka H, Nakayama M, Shimada A, Kitamura S, Hase Y, Narumi I, Tanaka A (2010) Comparative analysis of floral pigmentation between wild-type and white-flowered varieties of *Cyclamen graecum*. *J Hortic Sci Biotechnol* 85: 437-443.

Akita Y, Kitamura S, Hase Y, Narumi I, et al. (2011) Isolation and characterization of the fragrant cyclamen *O*-methyltransferase involved in flower coloration. *Planta* 234: 1127-1136

Asen S, Stewart RN, Norris KH (1972) Co-pigmentation of anthocyanins in plant tissues and its effect on color. *Phytochemistry* 3:1139-1144

Bieza K, Lois R (2001) An Arabidopsis mutant tolerant to lethal ultraviolet-B levels shows constitutively elevated accumulation of flavonoids and other phenolics. *Plant Physiol* 126:1105-1115

Boase MR, Lewis DH, Davies KM, Marshall GB, Patel D, Schwinn KE, Deroles SC (2010) Isolation and antisense suppression of *flavonoid 3',5'-hydroxylase* modifies flower pigments and colour in cyclamen. *BMC Plant Biol* 10:107

Boddu J, Svabek C, Sekhon R, Gevens A, Nicholson RL, Jones AD, Pedersen JF, Gustine DL, Chopra S (2004) Expression of a putative *flavonoid 3'-hydroxylase* in sorghum mesocotyls synthesizing 3-deoxyanthocyanidin phytoalexins. *Physiological and Molecular Plant Pathology* 2: 101-113

Bowles D, Lim EK, Poppenberger B, Vaistij FE (2006) Glycosyltransferases of lipophilic small molecules. *Annual Review of Plant Biology* 57: 567-597

Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72: 248-254

Brazier-Hicks M, Evans KM, Gershater MC, Puschmann H, Steel PG, Edwards R (2009) The *C*-glycosylation of flavonoids in cereals. *Journal of Biological Chemistry* 27: 17926-17934

Brugliera F, Barri-Rewell G, Holton TA, Mason JG (1999) Isolation and characterization of a flavonoid 3'-hydroxylase cDNA clone corresponding to the *Ht1* locus of *Petunia hybrida*. *Plant J* 19:441-451

Burbulis IE and Winkel-Shirley B (1999) Interactions among enzymes of the Arabidopsis flavonoid biosynthetic pathway. *Proc Natl Acad Sci USA* 96:12929-12934

Castellarin SD, Di Gaspero G, Marconi R, Nonis A, Peterlunger E, Paillard S, Adam-Blondon AF, Testolin R (2006) Colour variation in red grapevines (*Vitis vinifera* L.): genomic organisation, expression of *flavonoid 3'-hydroxylase*, *flavonoid 3',5'-hydroxylase* genes and related metabolite profiling of red cyanidin-/blue delphinidin-based anthocyanins in berry skin. *BMC Genomics* 7:12

Caputi L, Malnoy M, Goremykin V, Nikiforova S, Martens S (2012) A genome-wide phylogenetic reconstruction of family 1 UDP-glycosyltransferases revealed the expansion of the family during the adaptation of plants to life on land. *Plant J* 69: 1030-1042

Chapple C (1998) Molecular-genetic analysis of plant cytochrome P450-dependent monooxygenases. *Annu Rev Plant Physiol Plant Mol Biol* 49:311-343

Cornea-Cipcigan M, Pamfil D, Sisea CR, Gavris CP, et al. (2019) A review on *Cyclamen* species: transcription factors vs. pharmacological effects. *Acta Poloniae Pharmaceutica-Drug Research* 6: 919-938

Combet C, Blanchet C, Geourjon C, Deleage G (2000) NPS@: network protein sequence analysis. *Trends Biochem Sci* 25: 147-150

Dixon RA, Steele CL (1999) Flavonoids and isoflavonoids: a gold mine for metabolic engineering. *Trends Plant Sci* 4:394-400

Dooner HK, Nelson OE (1977) Controlling element-induced alterations in UDP-glucose: flavonoid glucosyltransferase, the enzyme specified by the bronze locus in maize. *Proceedings of the National Academy of Sciences* 12: 5623-5627

Du H, Wu J, Ji KX, Zeng QY, et al., (2015) Methylation mediated by an anthocyanin *O*-methyltransferase, is involved in purple flower coloration in *Paeonia*, *Journal of Experimental Botany* 21: 6563-6577

Falcone Ferreyra ML, Rodriguez E, Casas MI, Labadie G, Grotewold E, Casati P (2013) Identification of a bifunctional maize C- and O-glucosyltransferase. *Journal of Biological Chemistry* 4: 31678-31688

Fukusaki E, Kawasaki K, Kajiyama S, et al. (2004) Flower color modulations of *Torenia hybrida* by downregulation of *chalcone synthase* genes with RNA interference. *Journal of Biotechnology* 3: 229-240

Forkmann, G. (1991) Flavonoids as flower pigments: The formation of the natural spectrum and its extension by genetic engineering. *Plant Breed* 106: 1-26

Forkmann G, Heller W (1999) Biosynthesis of flavonoids. In: U. Sankawa (ed.): Comprehensive Natural Products Chemistry, Elsevier, Amsterdam, The Netherlands. 1:714-748

Forkmann G and Ruhnau B (1987) Distinct substrate specificity of dihydroflavonol 4-reductase from flowers of *Petunia hybrida*. *Zeitschrift für Naturforschung C* 42(9-10): 1146-1148

Geourjon C, Deléage G (1995) SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Bioinformatics* 6: 681-684

Grey-Wilson C (2002) Cyclamen: a guide for gardeners, Horticulturists and Botanists. Timber Press, England

Grotewold E (2006a) The genetics and biochemistry of floral pigments. *Annu. Rev. Plant Biol.* 57: 761-780

Han Y, Vimolmangkang S, Soria-Guerra R, Rosales-Mendoza S, Zheng D, et al. (2010) Ectopic expression of appl*e F3'H* genes contributes to anthocyanin accumulation in the Arabidopsis *tt7* mutant grown under nitrogen stress. *Plant Physiology* 2: 806-820

Hase Y, Akita Y, Kitamura S, Narumi I, Tanaka A (2012) Development of an efficient mutagenesis technique using ion beams: toward more controlled mutation breeding. *Plant Biotechnol* 29: 193-200

Hasemann CA, Kurumbail RG, Boddupalli SS, Peterson JA, Deisenhofer J (1995) Structure and function of cytochrome P450: a comparative analysis of three crystal structures. *Structure* 3:41-62

He F, Chen WK, Yu KJ, Ji XN, Duan CQ, Reeves MJ, Wang J (2015) Mole-cular and biochemical characterization of the UDP-glucose: anthocyanin 5-*O*-glucosyltransferase from *Vitis amurensis*. *Phytochemistry* 117: 363-372

Hichri I, Barrieu F, Bogs J, Kappel C, Delrot S, Lauvergeat V (2011) Recent advances in the transcriptional regulation of the flavonoid biosynthetic pathway. *Journal of Experimental Botany* 8: 2465-2483

Honda T and Saito N (2002) Recent progress in the chemistry of polyacylated anthocyanins as flower color pigment. *Heterocycles* 56: 633-692

Holton TA, Cornish EC (1995) Genetics and Biochemistry of Anthocyanin Biosynthesis. *Plant Cell* 7:1071-1083

Holton TA, and Tanaka Y (1994) Blue roses-A pigment of our imagination? *Trends Biotechnol* 12: 40-42

Hoshino A, Morita Y, Choi JD, Saito N, Toki T, Tanaka Y, Iida S (2003) Spontaneous mutations of the *flavonoid 3'-hydroxylase* gene in the three morning glory species displaying reddish flowers. *Plant Cell Physiol* 44:990-1001

Huits HS, Gerats AG, Kreike MM, Mol JN, Koes RE (1994) Genetic control of *dihydroflavonol 4-reductase* gene expression in *Petunia hybrida*. *Plant Journal* 6:295-310

Imayama T, Yoshihara N, Fukuchi-Mizutani M, Tanaka Y, Ino I, Yabuya T (2004) Isolation and characterization of a cDNA clone of UDP-glucose: anthocyanin 5-*O*-glucosyltransferase in *Iris hollandica*. *Plant Science* 6: 1243-1248

Ishizaka H, Uematsu J (1992) Production of interspecific hybrids of *Cyclamen persicum* Mill. and *C. hederifolium* Aiton by ovule culture. *Jpn J Breed* 42:353-366

Ishizaka H, Uematsu J (1995 b) Interspecific hybrids of *Cyclamen persicum* Mill. and *C. purpurascens* Mill. produced by ovule culture. *Euphytica* 82:31-37

Ishiguro K, Taniguchi M, Tanaka Y (2012) Functional analysis of *Antirrhinum kelloggii flavonoid 3'-hydroxylase* and *flavonoid 3',5'-hydroxylase* genes; critical role in flower color and evolution in the genus *Antirrhinum*. *J Plant Res* 125: 451-456

Ishizaka H (2018) Breeding of fragrant cyclamen by interspecific hybridization and ion-beam irradiation. *Breeding Science* 1: 25-34

Ishizaka H, Kondo E, Kameari N (2012) Production of novel flower color mutants from the fragrant cyclamen (*Cyclamen persicum*×*C. purpurascens*) by ion-beam irradiation. *Plant Biotechnol* 29: 201-208

Jeong ST, Goto-Yamamoto N, Hashizume K, Esaka M (2006) Expression of the *flavonoid 3'-hydroxylase* and *flavonoid 3',5'-hydroxylase* genes and flavonoid composition in grape (*Vitis vinifera*). *Plant Science*, 1:61-69

Jez JM, Bowman ME, Dixon RA, et al. (2000) Structure and mechanism of the evolutionarily unique plant enzyme chalcone isomerase. *Nat Struct Mol Biol* 9:786-791

Jiang WB, Yin QG, Wu RR, Zheng GS, Liu JY, Dixon RA, Pang YZ (2015) Role of a chalcone isomerase-like protein in flavonoid biosynthesis in *Arabidopsis thaliana. Journal of Experimental Botany* 22:7165-7179

Jin H, Martin C (1999) Multifunctionality and diversity within the plant *MYB*-gene family. *Plant Mol Biol* 41: 577-585

Johnson ET, Yi H, Shin B, Oh BJ, Cheong H, Choi G (1999) *Cymbidium hybrida* dihydroflavonol 4-reductase does not efficiently reduce dihydrokaempferol to produce orange pelargonidin-type anthocyanins. *Plant J* 19: 81-85.

Ju Z, Sun W, Meng X, Liang L, Li Y, Zhou T, Shen H, Gao X, Wang L (2018) Isolation and functional characterization of two 5-*O*-glucosyltransferases related to anthocyanin biosynthesis from *Freesia hybrida*. *Plant Cell Tiss Organ Cult* 135: 99-110

Kim S, Jones R, Yoo KS et al. (2004) Gold color in onions (*Allium cepa*): a natural mutation of the chalcone isomerase gene resulting in a premature stop codon. *Mol Genet Genomics* 272: 411-419

Kitamura S, Akita Y, Ishizaka H, Narumi I, Tanaka A (2012) Molecular characterization of an anthocyanin-related *glutathione S-transferase* gene in cyclamen. *Journal of Plant Physiology* 6: 636-642

Kitada C, Gong ZZ, Tanaka Y, Yamazaki M, Saito K (2001) Differential expression of two cytochrome P450s involved in the biosynthesis of flavones and anthocyanins in chemo-varietal forms of *Perilla frutescens*. *Plant and Cell Physiology* 12: 1338-1344

Koseki M, Goto K, Masuta C, et al. (2005) The star-type color pattern in *Petunia hybrida* 'Red Star' flowers is induced by sequence-specific degradation of Chalcone Synthase

RNA. *Plant and cell physiology* 46 (11): 1879-1883

Kogawa K, Kato N, Kazuma K, Noda N, Suzuki M (2007) Purification and characterization of UDP-glucose: anthocyanin 3',5'-*O*-glucosyltransferase from *Clitoria ternatea*. *Planta* 226 (6): 1501-1509

Kobayashi S, Ishimaru M, Ding CK, Yakushiji H, Goto N (2001) Comparison of UDP-glucose: *flavonoid 3*-O-*glucosyltransferase* (*UFGT*) gene sequences between white grapes (*Vitis vinifera*) and their sports with red skin. *Plant Science* 3: 543-550

Kubo A, Arai Y, Nagashima S, Yoshikawa T (2004) Alteration of sugar donor specificities of plant glycosyltransferases by a single point mutation. *Arch Biochem Biophys* 429: 198-203

Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol* 7:1870-1874

Li C, Qiu J, Yang G, Huang S, Yin J (2016) Isolation and characterization of a R2R3-MYB transcription factor gene related to anthocyanin biosynthesis in the spathes of *Anthurium andraeanum* (Hort.). *Plant Cell Rep* 35: 2151-2165

Liu H, Su B, Zhang H, Gong J, Zhang B, Liu Y, Du L (2019) Identification and functional analysis of a flavonol synthase gene from grape hyacinth. *Molecules* 24: 1579

Lim EK, Ashford DA, Hou B, Jackson RG, Bowles DJ (2004) Arabidopsis glycosyltransferases as biocatalysts in fermentation for regioselective synthesis of diverse quercetin glucosides. *Biotechnology and Bioengineering* 5: 623-631.

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time

quantitative PCR and the $2^{-\Delta\Delta CT}$ *Method. Methods* 25(4):402-408

Luo P, Ning G, Wang Z, Shen Y, Jin H, Li P, Huang S, Zhao J, Bao M (2016) Disequilibrium of flavonol synthase and dihydroflavonol-4-reductase expression associated tightly to white vs. red color flower formation in plants. *Front. Plant Sci* 6: 1257

Mackenzie PI, Bock KW, Burchell B, Guillemette C, Ikushiro S, Iyanagi T, Miners JO, Owens IS, Nebert DW (2005) Nomenclature update for the mammalian *UDP glycosyltransferase* (UGT) gene superfamily. *Pharmacogenet Genomics*15: 677-685

Masukawa T, Cheon K-S, Mizuta D, Nakatsuka A, Kobayashi N (2018) Insertion of a retrotransposon into a flavonoid 3'-hydroxylase homolog confers the red root character in the radish (*Raphanus sativus* L. var. longipinnatus L. H. Bailey). *Hort J* 87: 89-96.

Martin C, Prescott A, Mackay S, Bartlett J, Vrijlandt E (1991) Control of anthocyanin biosynthesis in flowers of *Antirrhinum majus*. *P1ant J* 1: 37-49

Martens S, Mithöfer A (2005) Flavones and flavone synthases, *Phytochemistry* 20: 2399-2407

Meech R, Hu DG, Mckinnon RA, Mubarokah SN, Haines AZ, Nair PC, Rowland A, Mackenzie PI (2019) The UDP-Glycosyltransferase (UGT) Superfamily: new members, new functions, and novel paradigms. *Physiological Reviews* 2: 1153-1222

Mol J, Cornish E, Mason J, Koes R (1999) Novel coloured flowers. *Curr. Opin. Biotech.* 10: 198-201

Morgret ML, Huang GH, Huang JK (2005). DNA sequence analysis of three clones containing chalcone synthase gene of *Petunia hybrida*. *FASEB J* 19:303-303

Mol J, Grotewold E, Koes R (1998) How genes paint flowers and seeds. *Trends in Plant Science* 6: 212-217

Murray MG, Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Research* 19: 4321-4326

Murakami K, Mihara K, Omura T (1994) The transmembrane region of microsomal cytochrome P450 identified as the endoplasmic reticulum retention signal. *J Biochem* 116:164-175

Nakatsuka T, Sato K, Takahashi H, Yamamura S, Nishihara M (2008) Cloning and characterization of the UDP-glucose: anthocyanin 5-*O*-glucosyltransferase gene from blue-flowered gentian. *J Exp Bot* 59: 1241-1252

Nakamura N, Katsumoto Y, Brugliera F, Demelis L, et al., (2015) Flower color modification in *Rosa hybrida* by expressing the *S*-adenosylmethionine: *anthocyanin 3',5'*-O-*methyltransferase* gene from *Torenia hybrida*. *Plant Biotechnology* 2:109-117

Nakatsuka A, Mizuta D, Kii Y, Miyajima I, Kobayashi N (2008) Isolation and expression analysis of flavonoid biosynthesis genes in evergreen azalea. *Scientia Horticulturae*, 4: 314-320

Nakatsuka T, Nishihara M, Mishiba K, et al. (2005) Two different mutations are involved in the formation of white-flowered gentian plants. *Plant Science* 5:949-958

Nicholson RL, Schmidt RH (1992) Phenolic compounds and their role in disease resistance. *Annu Rev Phytopathol* 30:369-389

Nishihara M, Nakatsuka T, Yamamura S. (2005) Flavonoid components and flower color

change in transgenic tobacco plants by suppression of *chalcone isomerase* gene. *FEBS Lett* 579:6074-6078

Nishihara M, Yamada E, Saito M et al. (2014) Molecular characterization of mutations in white-flowered torenia plants. *BMC Plant Biol* 14: 86

Nitarska D, Stefanini C, Haselmair-Gosch C, et al. (2018). The rare orange-red colored *Euphorbia pulcherrima* cultivar 'Harvest Orange' shows a nonsense mutation in a *flavonoid 3'-hydroxylase* allele expressed in the bracts. *BMC Plant Biol* 18:216

Noda N, Kanno Y, Kato N, Kazuma K, Suzuki M (2004) Regulation of gene expression involved in flavonol and anthocyanin biosynthesis during petal development in lisianthus (*Eustoma grandiflorum*). *Physiol Plant* 122: 305-313

Noda KI, Glover BJ, Linstead P, Martin C (1994) Flower color intensity depends on specialized cell shape controlled by a Myb-related transcription factor. *Nature* 369: 661-664

Noguchi A, Sasaki N, Nakao M, Fukami H, Takahashi S, Nishino T, Nakayama T (2008) cDNA cloning of glycosyltransferases from Chinese wolfberry (*Lycium barbarum* L.) fruits and enzymatic synthesis of a catechin glucoside using a recombinant enzyme (UGT73A10). *Journal of Molecular Catalysis B: Enzymatic* 55 (1-2): 84-92

Nuraini L, Ando Y, Kawai K, et al. (2020) Anthocyanin regulatory and structural genes associated with violet flower color of *Matthiola incana*. *Planta* 251, 61

Ogata Jun, Sakamoto T, Yamaguchi M, Kawanobu S, Yoshitama K (2001) Isolation and characterization of anthocyanin 5-*O*-glucosyltransferase from flowers of *Dahlia variabilis*. *J Plant Physiol* 158: 709-714

Okitsu N, Mizuno T, Matsui K, Choi SH, Tanaka Y (2018) Molecular cloning of flavonoid biosynthetic genes and biochemical characterization of anthocyanin *O*-methyltransferase of *Nemophila menziesii* Hook. and Arn. *Plant Biotechnology* 1: 9-16

Olsen KM, Hehn A, Jugdé H, Slimestad R, Larbat R, Bourgaud F, Lillo C (2010) Identification and characterization of CYP75A31, a new flavonoid 3',5'-hydroxylase isolated from *Solanum lycopersicum*. *BMC Plant Biology* 10 (1):21

Park S, Kim DH, Park BR, Lee JY, Lim SH (2019) Molecular and functional characterization of *Oryza sativa* flavonolsynthase (OsFLS), a bifunctional dioxygenase. *Journal of Agricultural and Food Chemistry* 26:7399-7409

Peer WA, Brown DE, Tague BW, Muday GK, Taiz L, Murphy AS (2001) Flavonoid accumulation patterns of *Transparent Testa* mutants of Arabidopsis. *Plant Physiol* 126: 536-548

Pichersky E, Gang DR (2000) Genetics and biochemistry of secondary metabolites in plants: an evolutionary perspective. *Trends Plant Sci* 5: 439-445

Reimold U, Kroeger M, Kreuzaler F, Hahlbrock K (1983). Coding and 3' non-coding nucleotide sequence of chalcone synthase messenger RNA and assignment of amino acid sequence of the enzyme. *EMBO J* 2: 1801-1806.

Rosati C, Simoneau P, Treutter D, et al. (2003) Engineering of flower color in forsythia by expression of two independently-transformed *dihydroflavonol 4-reductase* and *anthocyanidin synthase* genes of flavonoid pathway. *Molecular breeding* 12 (3): 197-208

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406-425

Saslowsky D and Winkel-Shirley B (2001) Localization of flavonoid enzymes in Arabidopsis roots. *Plant J* 1: 37-48

Saslowsky DE, Dana CD, Winkel-shirley B (2000) An allelic series for the chalcone synthase locus in *Arabidopsis*. *Gene* 255(2):127-138

Schoenbohm C, Martens S, Eder C, Forkmann G, Weisshaar B (2000) Identification of the *Arabidopsis thaliana flavonoid 3'-hydroxylase* gene and functional expression of the encoded P450 enzyme. *Biol Chem* 381:749-753

Shao H, He X, Achnine L, Blount JW, Dixon RA, Wang X (2005) Crystal structures of a multifunctional triterpene/flavonoid glycosyltransferase from *Medicago truncatula*. *J.Plant Cell* 17: 3141-3154

Shih CH, Chu IK, Yip WK, Lo C (2006) Differential expression of two *flavonoid 3'-hydroxylase* cDNAs involved in biosynthesis of anthocyanin pigments and 3-deoxyanthocyanidin phytoalexins in Sorghum. *Plant Cell Physiol.* 47(10): 1412-1419

Sigurdson GT, Robbins RJ, Collins TM, Giusti MM (2017) Spectral and colorimetric characteristics of metal chelates of acylated cyanidin derivatives. *Food Chemistry* 221: 1088-1095

Stafford HA (1974) Possible multi-enzyme complexes regulating the formation of C6-C3 phenolic compounds and lignins in higher plants. *Rec. A dv. Phytochem.* 8:53-79

Strack D, Vogt T, Schliemann W (2003) Recent advances in betalain research.

Phytochemistry 62: 247-269

Sui X, Zhao M, Xu Z, Zhao L, Han X (2018) *RrGT2*, a key gene associated with anthocyanin biosynthesis in *Rosa rugosa*, was identified via virus-induced gene silencing and overexpression. *Int. J. Mol. Sci*. 19: 4057

Sun W, Liang L, Meng X, Li Y, Gao F, Liu X, Wang S, Gao X, Wang L (2016) Biochemical and molecular characterization of a flavonoid 3-*O*-glycosyltransferase responsible for anthocyanins and flavonols biosynthesis in *Freesia hybrida*. *Front Plant Sci* 7:410

Takamura T, Aizawa M, Kim SY, Nakayama M, Ishizaka H (2005). Inheritance of flower pigment in crosses between cyclamen cultivars and *Cyclamen purpurascens*. *Acta Hortic* 673: 437-441

Takatori Y, Shimizu K, Ogata J, Endo H, Ishimaru K, Okamoto S, Hashimoto F (2015) Cloning of the *flavonoid 3'-hydroxylase* gene of *Eustoma grandiflorum* (Raf.) Shinn. (*EgF3'H*) and complementation of an *F3'H*-deficient Mutant of *Ipomoea nil* (L.) Roth. by heterologous expression of *EgF3'H*. *Horic J* 84: 131-139

Tanaka Y, Brugliera F (2013) Flower color and cytochromes P450. *Philosophical Transactions of the Royal Society B: Biological Sciences* 368, 20120432

Tanaka Y, Sasaki N, Ohmiya A (2008) Biosynthesis of plant pigments: anthocyanins, betalains and carotenoids. *The Plant Journal* 4: 733-749

Thompson JD, Higgins DG, Gibson TJ (1994) Clustal-W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673-4680

Tsuda S, Fukui Y, Nakamura N, Katsumoto Y, et al. (2004) Flower color modification of *Petunia hybrida* commercial varieties by metabolic engineering. *Plant Biotechnology* 5: 377-386

Ueyama Y, Suzuki K, Fukuchi-Mizutani M, Fukui Y, Miyazaki K, Ohkawa H, Kusumi T, Tanaka Y (2002). Molecular and biochemical characterization of torenia *flavonoid 3'-hydroxylase* and *flavone synthase* II and modification of flower color by modulating the expression of these genes. *Plant Sci* 163:253-263

Vogt T, Jones P (2000) Glycosyltransferases in plant natural product synthesis: characterization of a supergene family. *Trends Plant Sci* 5: 380-386

Winkel-Shirley B (2001) Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant Physiol* 126: 485-493

Winkel-Shirley B (2002) Biosynthesis of flavonoids and effects of stress. *Curr. Opin. Plant Biol* 5:218-223

Winkler RG, Helentjaris T (1995) The maize *Dwarf3* gene encodes a cytochrome P450-mediated early step in gibberellin biosynthesis. *Plant Cell* 7:1307-1317

Wu Yq, Wang Tl, Xin Y, Wang Gb, Xu La (2020) Overexpression of the *GbF3'H1* gene enhanced the epigallocatechin, gallocatechin, and catechin contents in transgenic populus. *J. Agric. Food Chem.* 68, 4: 998-1006

Xu BB, Li JN, Zhang XK, Wang R, Xie LL, Chai YR (2007) Cloning and molecular characterization of a functional *flavonoid 3'-hydroxylase* gene from *Brassica napus*. *J Plant Physiol* 164:350-363

Yamazaki S, Sato K, Suhara K, Sakaguchi M, Mihara K, Omura T (1993) Importance of the proline-rich region following signal-anchor sequence in the formation of correct conformation of microsomal cytochrome P450s. *J Biochem* 114:652-657

Yamazaki M, Gong Z, Fukuchi-Mizutani M, Fukui Y, Tanaka Y, Kusumi T, Saito K (1999) Molecular cloning and biochemical characterization of a novel anthocyanin 5-*O*-glucosyltransferase by mRNA differential display for plant forms regarding anthocyanin. *J Biol Chem* 274: 7405-7411

Yamazaki M, Yamagishi E, Gong Z, Fukuchi-Mizutani M, Fukui Y, Tanaka Y, Kusumi T, Yamaguchi M, Saito K (2002) Two flavonoid glucosyltransferases from *Petunia hybrida*: molecular cloning, biochemical properties and developmentally regulated expression. *Plant Mol Biol* 48: 401-411

Yamaguchi T, Fukada-Tanaka S, Inagaki Y, Saito N, et al. (2001) Genes encoding the vacuolar Na^+/H^+ exchanger and flower coloration. *Plant and Cell Physiology* 5: 451-461

Yoshida K, Mori M, Kondo T (2009) Blue flower color development by anthocyanins: from chemical structure to cell physiology. *Nat. Prod. Rep.* 26: 884-915

Yonekura-Sakakibara K, Hanada K (2011) An evolutionary view of functional diversity in family 1 glycosyltransferases. *Plant J* 66: 182-193

Yuan Y, Ma X, Tang D, Shi Y (2014) Comparison of anthocyanin components, expression of anthocyanin biosynthetic structural genes, and *TfF3'H1* sequences between *Tulipa fosteriana* 'Albert heijn' and its reddish sport. *Scientia Horticulturae* 175: 16-26

Zhang X, Xu Z, Yu X, Zhao L, Zhao M, Han X, Qi S (2019) Identification of two novel

R2R3-MYB transcription factors, *PsMYB114L* and *PsMYB12L*, related to anthocyanin biosynthesis in *Paeonia suffruticosa*. *Int. J. Mol. Sci.* 20:1055

Zhang YZ, Cheng YW, Xu SZ, Ma HP, Han JM, Zhang Y (2020) Tree peony variegated flowers show a small insertion in the *F3'H* gene of the acyanic flower parts. *BMC Plant Biology* 20:211

Zhao DQ, Tao J, Han CX, Ge JT (2012b). Flower color diversity revealed by differential expression of flavonoid biosynthetic genes and flavonoid accumulation in herbaceous peony (*Paeonia lactiflora* Pall.). *Mol. Biol. Rep.* 39:11263-11275

Zhou W, Gong Y, Lu X, Huang C, Gao F (2012) Molecular cloning and characterization of a *flavonoid 3'-hydroxylase* gene from purple-fleshed sweet potato (*Ipomoea batatas*). *Mol Biol Rep* 39:295-302

Related publications

Journal Article:

1. Kang X.F., Akita Y., Mikami R. Isolation and analysis of *flavonoid 3'-hydroxylase* (*F3'H*) genes from *Cyclamen* 'Strauss'. *Acta Horticulturae*

2. Kang X.F., Mikami R., Akita Y. Characterization of 5-*O*-glucosyltransferase involved in anthocyanin biosynthesis in *Cyclamen purpurascens*. *Plant Biotechnology*

3. Wu H.S., Noda N., Mikami R., Kang X.F., Akita Y. Insertion of a novel transposable element disrupts the function of an anthocyanin biosynthesis-related gene in *Echinacea purpurea*'. *Scientia Horticulturae*

Conferences:

1. 康暁飛，野田尚信，秋田祐介，シクラメン *Flavonoid 3'-Hydroxylase* (*F3'H*) 遺伝子の単離と解析. 園芸学会（2019年度）秋季大会. *14 to 17 September 2019, Shimane, Japan.*

2. Kang X.F., Akita Y., Identification and Characterization of Genes Responsible for Flower Coloration in Cyclamens. *The 3rd Asian Horticultural Congress 2020 (AHC2020). 15 to 17 December 2020, Bangkok, Thailand. (on line)*

Acknowledgements

Time flies, and three years are fleeting. Although I cannot speak Japanese fluently, I still have gained a lot of more precious things.

I would like to extend my sincere gratitude to Professor Hatada my supervisor, for his guidance and encouragement.

I would like to express my heartfelt gratitude to Dr. Akita, who taught me not only how to do it, but also how to think. He is my teacher and best friend all my life.

Thanks to Miss Parisa, Miss Mikami, Miss Morimura and everyone in Akita laboratory for their help in my study and life. They are all my best Japanese teachers.

I would like to express my deeply indebted to Mr Ju and Mrs Ju, for giving me the warmth of home in a foreign country.

I would like to thanks Qin Xiyan, Tai Zhongxu, Lv Rui, Li Xusheng, Yu Yimin, for their company when I was happy and sad. Every day they spent with me is worth remembering.

Finally, special thanks should go to my dear family, their dedication and support are the motivation for my progress.